

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 09:23:41 ; Search time 38.6861 Seconds
(without alignments)
287.115 Million cell updates/sec

Title: US-09-894-030-3
Perfect score: 545
Sequence: 1 MEKLFKEVKLEELENOKGG.....DDFDLDVVKVSKODSKITPQ 100

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 segs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

```

Database :
1: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT: *
2: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT: *
3: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT: *
4: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT: *
5: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT: *
6: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT: *
7: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT: *
8: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT: *
9: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT: *
10: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT: *
11: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT: *
12: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT: *
13: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT: *
14: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT: *
15: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT: *
16: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT: *
17: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT: *
18: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT: *
19: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT: *
20: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT: *
21: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT: *
22: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT: *

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query			Description	
	Score	Match	Length	ID	Description
1	304	55.8	56	AAV03205	Amino acid sequenc
2	211	38.7	37	AAV03207	Amino acid sequenc
3	202	37.1	37	AAV03218	Amino acid sequenc
4	142.5	26.1	576	AAV12229	TrpE/androgen rece
5	142.5	26.1	918	AAV12223	Human androgen rec
6	142.5	26.1	918	AAV33491	Human androgen rec
7	142.5	26.1	919	AAV03109	Human androgen rec
8	142.5	26.1	919	AAV09096	Human androgen rec
9	142.5	26.1	919	AAV14793	Androgen receptor
10	142.5	26.1	919	AAV178914	Human androgen rec
11	140	25.7	94	AAV71269	Human gene 1-encod

12	135.5	24.9	240	16	AAR78738	Murine mvl protein
13	135.5	24.9	240	18	AAR78738	Murine BMP-12 homo
14	135.5	24.9	240	22	AAE10594	Murine partial mvl
15	134.5	24.7	101	22	ABB27649	Human peptide #300
16	134.5	24.7	101	22	ABB32819	Peptide #325 enco
17	134.5	24.7	101	22	ABB318301	Human #300 enco
18	134.5	24.7	101	22	AAM363626	Human brain expres
19	134.5	24.7	101	22	AAM660005	Human bone marrow
20	134.5	24.7	101	22	AAM13874	Peptide #308 enco
21	134.5	24.7	101	22	AAM36281	Peptide #318 enco
22	134.5	24.7	101	22	AAM01617	Peptide #299 enco
23	129.5	23.8	738	19	AAM56163	New DNA sequence i
24	129	23.7	268	18	AAW1640	Human calpain smal
25	129	23.7	268	22	AAB86127	Human calpain 30KD
26	127.5	23.4	118	22	ABB58597	Drosophila melanog
27	127	23.3	399	22	ABB60010	Drosophila melanog
28	126.5	23.2	348	22	ABB59015	Drosophila melanog
29	126	23.1	258	21	AAG51723	Arabidopsis thalia
30	126	23.1	307	18	AAW18563	Novel fusion prote
31	125	22.9	50	18	AAW13633	Apolipoprotein (a)
32	125	22.9	161	16	AAG65182	GDF-7 C-terminal r
33	125	22.9	569	22	AAM79339	Human protein SEQ
34	125	22.9	569	22	AAR79340	Human protein SEQ
35	124.5	22.8	321	15	AAR54683	Human fibrillarin,
36	124.5	22.8	324	21	AAB43918	Human cancer assoc
37	124.5	22.8	357	22	ABB29912	Peptide #2563 enco
38	124.5	22.8	357	22	ABB35090	Peptide #2596 enco
39	124.5	22.8	357	22	ABB20509	Protein #2508 enco
40	124.5	22.8	357	22	AAM55912	Human brain expres
41	124.5	22.8	357	22	AAM62822	Human bone marrow
42	124.5	22.8	357	22	AAM16105	Peptide #2539 enco
43	124.5	22.8	357	22	AAM28596	Peptide #2633 enco
44	124.5	22.8	357	22	AAW03832	Peptide #2514 enco
45	124	22.8	50	18	AAW13632	Apolipoprotein (a)

ALIGNMENTS

RESULT 1

AAV03205
ID AAV03205 standard: Protein: 56 AA.

XX
AC: AAY03205:

XX DT 03-AUG-1999 (first entry)

XX
DE Amino acid sequence of sunA protein.

Sublancin 168; antimicrobial; food preservative; pro-sublancin 168;
 KW Gram-positive bacteria; pre-sublancin 168; sunA; sunT
 XX Gram-positive bacteria; pre-sublancin 168; sunA; sunT

XX
05
Bacillus subtilis.

XX PN WO9903352-A1.

XX
28-JAN-1999
PD

XX 17-JUL-1998: 98WO-IIS14547.

XX 18-III-1997: 97US-0053035.
PR

XX
PA (UJMA-) UNIV MARYLAND BALTIMORE

XX PT Hansen .TN:

XX
DB
WPT: 1999-131752/11.

DR N-PSDB; AAX28031.
XX

pt New antimicrobial peptide, sublancin 100, from *Bacillus subtilis*
 pt used for, e.g., treatment of infections caused by Gram negative
 pt bacteria and as food preservative

XX

Claim 6; Fig 3; 71pp; English.

This is the amino acid sequence of the sunA protein encoded by the prosublancin 168 nucleotide sequence used in the method of the invention. The peptide designated subslancin 168, is an antimicrobial useful for treating infections and preserving food against spoilage bacteria, particularly Gram-positive bacteria. Pro-sublancin 168 and pre-sublancin 168, are the precursors of subslancin 168. Sublancin 168 is very stable at low pH and can be autoclaved without damage. It does not decompose after 2 years in aqueous solution of about neutral pH.

Sequence 56 AA;

Query Match 55.8%; Score 304; DB 20; Length 56;
Best Local Similarity 100.0%; Pred. No. 1.2e-24;
Matches 56; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MEKLFKEVKLEENQKSGGLGKQAALWLQCAAGGTIGCGGAVACQNYRQFCR 56
1 MEKLFKEVKLEENQKSGGLGKQAALWLQCAAGGTIGCGGAVACQNYRQFCR 56

RESULT 2

AY03207
ID AAY03207 standard; Protein; 37 AA.

AY03207;

03-AUG-1999 (first entry)

Amino acid sequence of prosublancin 168.

Sublancin 168; antimicrobial; food preservative; pro-sublancin 168;
Gram-positive bacteria; pre-sublancin 168.

Bacillus subtilis.

WO9903352-A1.

28-JAN-1999.

17-JUL-1998; 98WO-US14547.

18-JUL-1997; 97US-0053035.

(UYMA-) UNIV MARYLAND BALTIMORE.

Hansen JN;

WPI; 1999-131752/11.

New antimicrobial peptide, subslancin 168, from Bacillus subtilis -
used for, e.g. treatment of infections caused by Gram negative
bacteria and as food preservative

Claim 1; Pages 52-53; 71pp; English.

This is the amino acid sequence of prosublancin 168 used in the method of the invention. The peptide designated subslancin 168, is an antimicrobial useful for treating infections and preserving food against spoilage bacteria, particularly Gram-positive bacteria. Pro-sublancin 168 and pre-sublancin 168, are the precursors of subslancin 168. Sublancin 168 is very stable at low pH and can be autoclaved without damage. It does not decompose after 2 years in aqueous solution of about neutral pH.

Sequence 37 AA;

Query Match 38.7%; Score 211; DB 20; Length 37;
Best Local Similarity 100.0%; Pred. No. 3.9e-15;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MEKLFKEVKLEENQKSGGLGKQAALWLQCAAGGTIGCGGAVACQNYRQFCR 56

OY 20 GLGKQAALWLQCAAGGTIGCGGAVACQNYRQFCR 56
DB 1 GLGKQAALWLQCAAGGTIGCGGAVACQNYRQFCR 37

RESULT 3

AY03218
ID AAY03218 standard; Protein; 37 AA.

AY03218;

03-AUG-1999 (first entry)

Amino acid sequence of subslancin 168.

Sublancin 168; antimicrobial; food preservative; pro-sublancin 168;
Gram-positive bacteria; Pro-sublancin 168; pre-sublancin 168.

Bacillus subtilis.

Key Location/Qualifiers

Disulfide-bond 7..36

Disulfide-bond 14..29

Modified-site 16

/note= "dehydrogenated Ser (Dha)"

Modified-site 19

/note= "dehydrogenated Thr (Dhb)"

Cross-links 19..22

/note= "thioether bridge"

WO9903352-A1.

28-JAN-1999.

17-JUL-1998; 98WO-US14547.

18-JUL-1997; 97US-0053035.

(UYMA-) UNIV MARYLAND BALTIMORE.

Hansen JN;

WPI; 1999-131752/11.

New antimicrobial peptide, subslancin 168, from Bacillus subtilis -
used for, e.g. treatment of infections caused by Gram negative
bacteria and as food preservative

Claim 13; Fig 7; 71pp; English.

This is the amino acid sequence of subslancin 168 used in the method of the invention. The peptide designated subslancin 168, is an antimicrobial useful for treating infections and preserving food against spoilage bacteria, particularly Gram-positive bacteria. Pro-sublancin 168 and pre-sublancin 168, are the precursors of subslancin 168. Sublancin 168 is very stable at low pH and can be autoclaved without damage. It does not decompose after 2 years in aqueous solution of about neutral pH.

Sequence 37 AA;

Query Match 37.1%; Score 202; DB 20; Length 37;
Best Local Similarity 94.6%; Pred. No. 3.4e-14;
Matches 35; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 20 GLGKQAALWLQCAAGGTIGCGGAVACQNYRQFCR 56
DB 1 GLGKQAALWLQCAAGGTIGCGGAVACQNYRQFCR 37

RESULT 4

AAR12229
ID AAR12229 standard; Protein; 576 AA.

```

XX AC AAR12229;
XX 20-AUG-1991 (first entry)
XX TrpE/androgen receptor N-terminal domain fusion protein.
XX androgen receptor; AR; DNA-binding protein; steroid hormone.
XX Homo sapiens.
XX Key Location/Qualifiers
XX Protein 1..323
XX Region 324..329 /label= 33kD trpE protein
XX Protein 330..571 /label= 6 amino acid linker
XX Region 572..576 /label= N-terminal region of AR
XX /label= 5 amino acid linker
XX WO9107423-A.
XX 30-MAY-1991.
XX 19-OCT-1990; 90WO-US06015.
XX 17-NOV-1989; 89US-0438775.
XX (ARCH-) ARCH DEV CORP.
XX Liao S, Chang C;
XX WPI; 1991-178048/24.
XX N-PSDB; AAQ12007.
XX Androgen receptor and TR2 DNA binding proteins - DNA sequences
XX and antibodies for detection and quantification methods
XX Example 13; Fig 9; 79pp; English.
XX To express an androgen receptor fusion protein in E.coli, the PATH
XX expression system was used. The trpE is insoluble so partially
XX purified induced fusion protein is obtained by simply lysing the
XX E.coli and precipitating the insoluble fusion protein. The fusion
XX protein was used for immunisation to obtain monoclonal anti-AR
XX antibodies.
XX Sequence 576 AA;
XX
XX Query Match 26.1%; Score 142.5; DB 12; Length 576;
XX Best Local Similarity 50.8%; Pred. No. 9.9e-07;
XX Matches 32; Conservative 3; Mismatches 13; Indels 15; Gaps 3;
XX
XX QY 18 GSGLGKAOCAALW--LQCAAGGTI--GCGGAVACQNYRQFCRGGGGGGGGGGGGG 73
XX ||| | : | | : |||| | | | | | | | | | | | | | | | | | | | |
XX Db 416 GSGSPSAASSSWHTLFTAEGQLYGPCGGG-----GGGGGGGGGGGGGGG 464
XX
XX QY 74 GGG 76
XX |||
XX Db 465 GGG 467
XX
XX RESULT 5
XX AAR12223
XX ID AAR12223 standard; Protein; 918 AA.
XX AC AAR12223;
XX 20-AUG-1991 (first entry)
XX Human androgen receptor.
XX
XX Query Match 26.1%; Score 142.5; DB 12; Length 576;
XX Best Local Similarity 50.8%; Pred. No. 9.9e-07;
XX Matches 32; Conservative 3; Mismatches 13; Indels 15; Gaps 3;
XX
XX QY 18 GSGLGKAOCAALW--LQCAAGGTI--GCGGAVACQNYRQFCRGGGGGGGGGGGGG 73
XX ||| | : | | : |||| | | | | | | | | | | | | | | | | | | | |
XX Db 416 GSGSPSAASSSWHTLFTAEGQLYGPCGGG-----GGGGGGGGGGGGGGG 464
XX
XX QY 74 GGG 76
XX |||
XX Db 465 GGG 467
XX
XX RESULT 5
XX AAR12223
XX ID AAR12223 standard; Protein; 918 AA.
XX AC AAR12223;
XX 20-AUG-1991 (first entry)
XX Human androgen receptor.
XX

```

```

KW hAR; DNA-binding protein; steroid hormone.
XX Homo sapiens.
XX Key Location/Qualifiers
XX Domain 556..626 /label= DNA-binding domain
XX /note= "cysteine-rich"
XX WO9107423-A.
XX 30-MAY-1991.
XX 19-OCT-1990; 90WO-US06015.
XX 17-NOV-1989; 89US-0438775.
XX (ARCH-) ARCH DEV CORP.
XX Liao S, Chang C;
XX WPI; 1991-178048/24.
XX N-PSDB; AAQ12001.
XX Androgen receptor and TR2 DNA binding proteins - DNA sequences
XX and antibodies for detection and quantification methods
XX Claim 25; Fig 3; 79pp; English.
XX
XX This sequence was deduced from a cDNA clone isolated by screening
XX commercially available human testis and prostate lambda gt11 cDNA
XX libraries. The sequence is very similar to that of rat AR and in
XX the DNA-binding domain it is identical to that of rat AR DNA-binding
XX domain. Homology comparisons with other known steroid receptors
XX indicate that hAR is more closely related to glucocorticoid,
XX mineralo-corticoid and progesterone receptors than to v-erb-A or to
XX receptors for oestrogen, vitamin D and thyroid hormones.
XX Sequence 918 AA;
XX
XX Query Match 26.1%; Score 142.5; DB 12; Length 918;
XX Best Local Similarity 50.8%; Pred. No. 1.6e-06;
XX Matches 32; Conservative 3; Mismatches 13; Indels 15; Gaps 3;
XX
XX QY 18 GSGLGKAOCAALW--LQCAAGGTI--GCGGAVACQNYRQFCRGGGGGGGGGGGGG 73
XX ||| | : | | : |||| | | | | | | | | | | | | | | | | | | | |
XX Db 417 GSGSPSAASSSWHTLFTAEGQLYGPCGGG-----GGGGGGGGGGGGGGG 465
XX
XX QY 74 GGG 76
XX |||
XX Db 466 GGG 468
XX
XX RESULT 6
XX AAY33491
XX ID AAY33491 standard; Protein; 918 AA.
XX AC AAY33491;
XX 19-JAN-2000 (first entry)
XX Human androgen receptor protein.
XX
XX Proapoptotic; dependence domain; p75NTR; androgen receptor; DCC;
XX huntingtin polypeptide; Machado-Joseph disease; SCAl; SCAl; SCAl;
XX atrophin-1; cell death; apoptosis; Huntington's disease; head trauma;
XX Alzheimer's disease; Kennedy's disease; spinocerebellar ataxia; stroke;
XX dentatorubropallidolysian atrophy; cell proliferation; cell survival;
XX neoplastic; malignant; autoimmune; fibrotic.
XX Homo sapiens.
XX WO9945944-A1.

```

XX PD 16-SEP-1999.

XX PF 11-MAR-1999; 99WO-US05250.

XX PR 12-MAR-1998; 98US-0041886.

XX PA (BURN-) BURNHAM INST.

XX PI Bredesen DE, Rabizadeh S;

XX DR WPI: 1999-561617/47.

XX DR N-PSDB; AA223424.

XX New proapoptotic dependence peptides, used to develop products for treating, e.g. Alzheimer's disease -

XX PT Disclosure; Page 90-93; 199pp; English.

XX This invention describes novel pure proapoptotic dependence peptides which comprise a sequence of an active dependence domain selected from dependence polypeptides consisting of p53NTR, androgen receptor, DCC, huntingtin polypeptide, Machado-Joseph disease gene product, SCAL, SCA2, SCA6 and atrophin-1 polypeptide. The proapoptotic peptides are capable of inducing cell death and can be used to develop products to mediate or inhibit apoptosis. The methods can be used for reducing the severity of a proapoptotic dependence domain mediated pathological conditions e.g. Huntington's disease, Alzheimer's disease, Kennedy's disease, Spinocerebellar ataxias, dentatorubropallidoluysian atrophy, Machado-Joseph disease, stroke or head trauma. They can also be used for reducing the severity of a pathological condition mediated by upregulated cell proliferation or cell survival e.g. neoplastic, malignant, autoimmune or fibrotic conditions. This sequence represents a human androgen receptor described in the method of the invention.

XX SQ Sequence 918 AA;

Query Match 26.1%; Score 142.5; DB 20; Length 918;

Best Local Similarity 42.5%; Pred. No. 1.6e-06;

Matches 37; Conservative 3; Mismatches 22; Indels 25; Gaps 4;

QY 12 ELENQKSGGLGKAQCAALWLQACSG-----GTIGCGGG---AVACQNYRQF----- 54

Db 382 KLENPLDYGSAAWAAAAA---QCRYGDLASLHGAGAGPGSGSPSAASSSWHTLFTAEAG 438

QY 55 -----CRGGGGGGGGGGGGGGGGGGGG 76

Db 439 QLYGPCGGGGGGGGGGGGGGGGGGGGGG 465

RESULT 7

AAP93109

ID AAP93109 standard; protein; 919 AA.

XX AC AAP93109;

XX DT 19-MAR-1990 (first entry)

XX DE Human androgen receptor.

XX KW Human androgen receptor; polyclonal antibody; cancer.

XX OS Homo sapiens.

XX PN WO8909791-A.

XX PD 19-OCT-1989.

XX PF 13-APR-1989; 89WO-US01548.

XX PR 14-APR-1988; 88US-0182646.

XX PA (JUNC-) UNIVERSITY OF NORTH CAROLINA.

XX PI French FS, Wilson EM, Joseph DR, Lubahn DB;

XX DR WPI: 1989-324206/44.

XX DR N-PSDB; AAN91772.

XX DNA encoding androgen receptor protein - useful for transforming eukaryotic hosts for protein expression and subsequent antibody prodn.

XX PS Disclosure; Fig. 4; 41pp; English.

XX Androgen receptor protein (AR) is used to produce mono- or poly-clonal antibodies. These are used for the detection and quantification of AR in the presence of endogenous androgen, as androgen will not interfere with binding. They may be used in assays to determine and quantify cellular distribution of AR in tumour tissue, and are esp. useful for evaluating prostate cancers to determine responsiveness to androgen withdrawal therapy.

XX SQ Sequence 919 AA;

Query Match 26.1%; Score 142.5; DB 10; Length 919;

Best Local Similarity 42.5%; Pred. No. 1.6e-06;

Matches 37; Conservative 3; Mismatches 22; Indels 25; Gaps 4;

QY 12 ELENQKSGGLGKAQCAALWLQACSG-----GTIGCGGG---AVACQNYRQF----- 54

Db 386 KLENPLDYGSAAWAAAAA---QCRYGDLASLHGAGAGPGSGSPSAASSSWHTLFTAEAG 442

QY 55 -----CRGGGGGGGGGGGGGGGGGGGG 76

Db 443 QLYGPCGGGGGGGGGGGGGGGGGGGGGG 469

RESULT 8

AAP90996

ID AAP90996 standard; protein; 919 AA.

XX AC AAP90996;

XX DT 28-FEB-1990 (first entry)

XX DE Human androgen receptor DNA clone.

XX KW Androgen receptor; TR2 polypeptide;

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

XX FT Region 1..919

XX FT /*tag= a

XX FT /product=98 kD polypeptide

XX FT Region 185..919

XX FT /*tag= b

XX FT /product=79 kD polypeptide

XX PN WO8909223-A.

XX PD 05-OCT-1989.

XX PF 24-MAR-1989; 89WO-US01238.

XX PR 30-MAR-1988; 88US-0176107.

XX PA (ARCH-) ARCH DEVELOPMENT CORP.

XX PI Liao S, Chang C;

XX DR WPI: 1989-309501/42.

XX DR N-PSDB; AAN91577.

XX New DNA encoding new androgen receptor and TR2 polypeptide(s) - able to bind DNA, and derived antibodies, useful for receptor assay and

QY	12	ELENQKSGSLGKAQCAALWLQCASG-----GTICGGG---AVACQNYRQF-----	54
Db	386	KLENPLDYGSAWAAA--QCRYGLASLHGAGAGPGSPSAASSWHTLTAEAG	442
QY	55	-----CRGGGGGGGGGGGGGGGGGG 76	
Db	443	QLYGPCGGGGGGGGGGGGGGGGGG 469	
RESULT 10			
AA	AY78914	AAAY78914 standard; protein; 919 AA.	
XX	AC	AAAY78914;	
XX	AC		
XX	DT	23-MAY-2000 (first entry)	
XX	DE	Human androgen receptor (AR) amino acid sequence.	
XX	KW	Androgen receptor; AR; androgen-independent activation; inhibitor;	
XX	KW	cancer; benign prostatic hyperplasia; hirsutism; androgenic alopecia;	
XX	KW	acne; breast cancer; Kennedy disease; prostate cancer.	
XX	OS	Homo sapiens.	
XX	XX	WO200001813-A2.	
XX	PN	13-JAN-2000.	
XX	PD		
XX	XX	30-JUN-1999; 99WO-CA00604.	
XX	PF		
XX	XX	30-JUN-1998; 98US-0091871.	
XX	PR	(UYBR-) UNIV BRITISH COLUMBIA.	
XX	XX		
XX	PA		
XX	XX	Sadar MD, Bruchovsky N, Gout PW, Snoek R, Mawji NR;	
XX	PI	WPI; 2000-182113/16.	
XX	DR		
XX	XX	Novel non-androgen ligand binding peptides for inhibiting	
XX	PT	androgen-independent activation of androgen receptor, used for	
XX	PT	screening compounds and for treatment of androgen-mediated diseases	
XX	PT	such as prostate cancer	
XX	XX	Disclosure; Page 7; 32pp; English.	
XX	PS		
XX	XX	This sequence represents the human androgen receptor (AR) amino acid	
CC	CC	sequence. The invention relates to a fragment of the AR corresponding to	
CC	CC	amino acids 234-391 (see AAY78913). The fragment mediates	
CC	CC	androgen-independent activation of the AR. The androgen receptor acts as	
CC	CC	a transcription factor, regulating the expression of certain	
CC	CC	androgen-responsive genes. Interaction of the AR with the protein kinase	
CC	CC	A signal transduction pathway involves interaction with the androgen	
CC	CC	independent region. The AR fragment and peptides derived from it can be	
CC	CC	used as agents for inhibiting androgen independent activation of the	
CC	CC	androgen receptor, as activation domains, and as a tool for screening	
CC	CC	for compounds which affect androgen-independent activation of the AR.	
CC	CC	The peptides, when used in combination with androgen deprivation,	
CC	CC	effectively limit androgen mediated disease progression. These diseases	
CC	CC	include cancer, benign prostatic hyperplasia, hirsutism, androgenic	
CC	CC	alopecia, acne, breast cancer, Kennedy disease, and especially prostate	
CC	CC	cancer. The peptides and nucleic acids encoding them, are especially used	
CC	CC	for the treatment of androgen-mediated diseases, especially prostate	
CC	CC	tumours in patients deprived of androgen.	
XX	XX		
SQ	Sequence	919 AA;	
Query Match 26.1%; Score 142.5; DB 21; Length 919;			
Best Local Similarity 42.5%; Pred. No. 1.6e-06;			
Matches 37; Conservative 3; Mismatches 22; Indels 25; Gaps 4			
QY	12	ELENQKSGSLGKAQCAALWLQCASG-----GTICGGG---AVACQNYRQF-----	54

```

Db 386 KLENPLDYGSAAAAA---OCRYGDLASLHGAGAGPGSGSPSAASSSWHLFTAEAG 442
QY 55 -----CRGGGGGGGGGGGGGGGGGG 76
Db 443 QLYGPGCGGGGGGGGGGGGGGGGGGG 469

RESULT 11
AAAG71269
ID AAG71269 standard; Protein; 94 AA.
XX
XX AAG71269;
AC
XX
DT 30-JUL-2001 (first entry)
XX
DE Human gene 1-encoded secreted protein HTEGI42, SEQ ID NO:118.
XX
KW Human; secreted protein; proliferative disorder; cancer; chromosome 15;
KW foetal abnormality; developmental abnormality; haematopoietic disorder;
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW inflammation; allergy; neurological disorder; Alzheimer's disease;
KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW cardiovascular disorder; angiogenic disorder; kidney disorder;
KW gastrointestinal disorder; pregnancy-related disorder; tumour;
KW endocrine disorder; infection; wound healing; vulnerability;
KW cell culture; chemotaxis; food additive;
KW binding partner identification.
XX
OS Homo sapiens.
XX
PN W0200132674-A1.
XX
PD 10-MAY-2001.
XX
XX 25-OCT-2000; 2000WO-US29360.
XX
XX 29-OCT-1999; 99US-0162211.
XX
XX 30-JUN-2000; 2000US-0215138.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ruben SM, Komatsoulis GA, Young PE, Moore PA;
XX
XX WPI: 2001-291051/30.
XX
XX N-PSDB; AAH31376.
XX
XX New nucleic acid molecule encoding a human secreted protein, useful for
XX preventing, treating or ameliorating medical conditions such as
XX rheumatoid arthritis, Alzheimer's disease and microbial infections -
XX
XX Claim 11; Page 496; 581pp; English.
XX
XX AAH31349-AAH31428 represent cDNAs corresponding to 26 human secreted
XX protein genes, and AAG71243-AAG71319 represent the proteins they encode.
XX AAG71320-AAG71403 represent human secreted protein fragments. The genes
XX and their corresponding secreted proteins are useful for preventing,
XX treating or ameliorating medical conditions, e.g., by protein or gene
XX therapy. Pathological conditions can be diagnosed by determining the
XX amount of the new protein in a sample or by determining the presence of
XX mutations in the new genes. Specific uses are described for each of the
XX 52 genes, based on the tissues in which they are most highly expressed,
XX and include developing products for the diagnosis or treatment of
XX proliferative disorders, cancer, tumours, foetal and developmental
XX abnormalities, haematopoietic disorders, diseases of the immune system,
XX AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
XX allergies, neurological disorders (e.g., Alzheimer's disease,
XX Parkinson's disease), cognitive disorders, schizophrenia, asthma,
XX skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
XX cardiovascular disorders, angiogenic disorders, kidney disorders,
XX gastrointestinal disorders, pregnancy-related disorders, endocrine
XX disorders, and infections. The proteins can also be used to aid wound

```

```

CC healing and epithelial cell proliferation, to prevent skin aging due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues, to identify their
CC cognate ligands or binding partners, and in chemotaxis, and can be used
CC as a food additive or preservative to modify storage properties.
CC Antibodies specific for a protein of the invention can be used in
CC alleviating symptoms associated with the disorders mentioned above, and
CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC immunosorbent assay (ELISA). The present sequence represents a human
CC secreted protein of the invention.
XX
XX Sequence 94 AA;

```

```

Query Match 25.7%; Score 140; DB 22; Length 94;
Best Local Similarity 59.6%; Pred. No. 2.7e-07;
Matches 28; Conservative 1; Mismatches 4; Indels 14; Gaps 2;

```

```

QY 33 CASGGTGGGGGAVACQNYRQFCRGGGGGGGGGGGGGGGMSK 79
Db 23 CQGGG-GGGGG-----GGGGGGGGGGGGGGGGGGVDK 55

```

RESULT 12

```

AAAR78738
ID AAR78738 standard; Protein; 240 AA.
XX
XX AAR78738;
AC
XX
XX 23-NOV-1995 (first entry)
XX
XX Murine mV1 protein.
XX
XX Bone morphogenetic protein; mV1; tendon; ligament.
XX
XX Mus musculus.
XX
XX Key Location/Qualifiers
XX Misc-difference 7 /label= V,A,E,G
XX Misc-difference 8 /label= S,P,T,A
XX Misc-difference 15 /label= S,R
XX Misc-difference 16 /label= L,P,Q,R
XX Misc-difference 26 /label= C,W
XX Misc-difference 37 /label= V,A,D,G
XX Misc-difference 103 /label= V,A,E,G
XX Misc-difference 104 /label= Q,K,E
XX
XX W09516035-A.
XX
XX 15-JUN-1995.
XX
XX 06-DEC-1994; 94WO-US14030.
XX
XX 02-NOV-1994; 94US-0333576.
XX
XX 07-DEC-1993; 93US-0164103.
XX
XX 23-MAR-1994; 94US-0217780.
XX
XX (GEM) GENETICS INST INC.
XX (HARD) HARVARD COLLEGE.
XX
XX Celeste AJ, Melton DA, Rosen VA, Thomsen GH, Wolfman NM;
XX Wozney JM;
XX
XX WPI: 1995-224320/29.
XX
XX N-PSDB; AAQ96223.
XX

```

PT Bone morphogenetic proteins -12 and -13 and corresp. DNA - used in
 XX compsn. for inducing tendon/ligament-like tissue formation
 PS Example; Page 68-70; 84pp; English.
 XX

CC Oligos #6 and #7 (AAQ96218 & AAQ96219) are used as primers for the
 CC amplification of a 275 bp DNA probe, the internal 269 bp of which
 CC corresp. to nts #607 to #865 of AAQ96207, from the BMP-12 encoding
 CC plasmid subclone pCR1-1#2. This probe was radioactively labelled
 CC and used to screen a murine genomic library. DNA sequence analysis
 CC of one of positively hybridising recombinants named MVR3 indicates
 CC that it encodes a portion of the mouse gene corresp. to the PCR
 CC product mV1 (murine homolog of the human BMP-12 sequence AAQ96207.
 CC The partial DNA sequence of this subclone and corresp. AA
 CC translation are given in AAQ96223 & AAR78738.
 XX

SQ Sequence 240 AA;

Query Match 24.9%; Score 135.5; DB 16; Length 240;
 Best Local Similarity 40.4%; Pred. No. 2.1e-06;
 Matches 36; Conservative 5; Mismatches 25; Indels 23; Gaps 4;

OY 2 EKLFKEVKLE-----ELEKQKSGLG--KAQCAALWLQ-CASGGTIGCGGAVACQN 50
 DB 53 ESLFREIPAQARALRAAAEPPDPGPGAGSRKANLGGRRRQRTALAGTRCXXGS----- 106

OY 51 YRQFCRGGGGGGGGGGGGGGGGGGGMSK 79
 DB 107 -----GGGGGGGGGGGGGGGGGGGAGR 129

RESULT 13
 AAW26594
 ID AAW26594 standard; Protein; 240 AA.

AC AAW26594;
 XX
 XX 21-JAN-1998 (first entry)

DE Murine BMP-12 homologue fragment.

KW BMP-12; bone morphogenetic protein; mouse; tendon; ligament;
 KW wound healing; tissue repair; carpal tunnel syndrome; tendonitis;
 KW therapy.

OS Mus musculus.

FH Key Location/Qualifiers

FT Misc-difference 7 /label= Val, Ala, Glu, Gly
 FT /note= "encoded by GNG"

FT Misc-difference 8 /label= Ser, Pro, Thr, Ala
 FT /note= "encoded by NCG"

FT Misc-difference 15 /label= Ser, Arg
 FT /note= "encoded by AGN"

FT Misc-difference 16 /label= Leu, Pro, Gln, Arg
 FT /note= "encoded by CNG"

FT Misc-difference 26 /label= Cys, Trp
 FT /note= "encoded by TGN"

FT Misc-difference 37 /label= Val, Ala, Asp, Gly
 FT /note= "encoded by GNC"

FT Misc-difference 103 /label= Val, Ala, Glu, Gly
 FT /note= "encoded by GNG"

FT Misc-difference 104 /label= Lys, Gln
 FT /note= "encoded by NAG"

XX

PN US5658882-A.
 XX 19-AUG-1997.

XX 07-DEC-1993; 93US-0164103.

XX 22-DEC-1994; 94US-0362670.

PR 07-DEC-1993; 93US-0164103.

PR 25-MAR-1994; 94US-0217780.

PR 02-NOV-1994; 94US-0333576.

XX (GEMY) GENETICS INST INC.

PA (HARD) HARVARD COLLEGE.

XX Celeste AJ, Melton DA, Rosen VA, Thomsen GH, Wolfman NM;

PI Wozney JM;

XX WPI; 1997-424270/39.

XX N-PSDB; AAT90396.

XX Inducing tendon and ligament formation using BMP-12, BMP-13 or MP-52

PT - useful for tissue healing and repair, treatment of tendonitis,

PT improving fixation of tendons to bone etc

XX Example 1; Column 65-68; 43pp; English.

XX This polypeptide comprises a fragment of a murine homologue of
 CC human bone morphogenetic protein 12 (BMP-12) (see also AAW26589).
 CC Its amino acid sequence was deduced from DNA subclone mV1 (see
 CC AAT90396), isolated from murine genomic DNA using primers (see
 CC AAT90393-94) based on human BMP-12 sequences. Human BMP-12, BMP-13
 CC (see AAW26591) and MP52 (see AAW26590) polypeptides are used in a
 CC claimed method for inducing tendon and ligament formation.

SQ Sequence 240 AA;

Query Match 24.9%; Score 135.5; DB 18; Length 240;
 Best Local Similarity 40.4%; Pred. No. 2.1e-06;
 Matches 36; Conservative 5; Mismatches 25; Indels 23; Gaps 4;

OY 2 EKLFKEVKLE-----ELEKQKSGLG--KAQCAALWLQ-CASGGTIGCGGAVACQN 50
 DB 53 ESLFREIPAQARALRAAAEPPDPGPGAGSRKANLGGRRRQRTALAGTRCXXGS----- 106

OY 51 YRQFCRGGGGGGGGGGGGGGGGGGGMSK 79

DB 107 -----GGGGGGGGGGGGGGGGGGGAGR 129

RESULT 14

AAE10984

ID AAE10984 standard; Protein; 240 AA.

XX AAE10984;

XX 18-DEC-2001 (first entry)

XX Murine partial mV1 protein.

XX Mouse; bone morphogenetic protein; BMP-12; vulnery; antiinflammatory;
 KW analgesic; gene therapy; transforming growth factor-beta; TGF-beta;
 KW tissue formation; wound healing; tissue repair; ligament defect;
 KW carpal tunnel syndrome; tendonitis; mV1.

XX Mus sp.

XX Key Location/Qualifiers

FT Misc-difference 7 /label= Unknown
 FT /note= "Encoded by GNG"

FT Misc-difference 8 /label= Unknown
 FT /note= "Encoded by NCG"

XX

us-09-894-030-3.rag

Thu Oct 24 14:39:59 2002

Qy 66 GGGGGGGGGGMSK 79
 |||||
Db 48 GGGGGGGGGRRRK 61

Search completed: October 24, 2002, 09:27:58
Job time : 41.6861 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 09:28:06 : Search time 152.555 Seconds
(without alignments)
230.724 Million cell updates/sec

Title: US-09-894-030-3

Perfect score: 545

Sequence: 1 MEKLFKEVKLEENQKSGG.....DDFDLVVVKSKQDKITPQ 100

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 3502263 seqs, 351980561 residues

Total number of hits satisfying chosen parameters: 3502263

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending Patents AA.Main:*

- 1: /cgn2_6/ptodata/1/paa/US086_COMB.pep.*
- 2: /cgn2_6/ptodata/1/paa/US06_COMB.pep.*
- 3: /cgn2_6/ptodata/1/paa/US07_COMB.pep.*
- 4: /cgn2_6/ptodata/1/paa/US080_COMB.pep.*
- 5: /cgn2_6/ptodata/1/paa/US081_COMB.pep.*
- 6: /cgn2_6/ptodata/1/paa/US082_COMB.pep.*
- 7: /cgn2_6/ptodata/1/paa/US083_COMB.pep.*
- 8: /cgn2_6/ptodata/1/paa/US084_COMB.pep.*
- 9: /cgn2_6/ptodata/1/paa/US085_COMB.pep.*
- 10: /cgn2_6/ptodata/1/paa/US086_COMB.pep.*
- 11: /cgn2_6/ptodata/1/paa/US087_COMB.pep.*
- 12: /cgn2_6/ptodata/1/paa/US088_COMB.pep.*
- 13: /cgn2_6/ptodata/1/paa/US089_COMB.pep.*
- 14: /cgn2_6/ptodata/1/paa/US090_COMB.pep.*
- 15: /cgn2_6/ptodata/1/paa/US091_COMB.pep.*
- 16: /cgn2_6/ptodata/1/paa/US092_COMB.pep.*
- 17: /cgn2_6/ptodata/1/paa/US093_COMB.pep.*
- 18: /cgn2_6/ptodata/1/paa/US094_COMB.pep.*
- 19: /cgn2_6/ptodata/1/paa/US095_COMB.pep.*
- 20: /cgn2_6/ptodata/1/paa/US096_COMB.pep.*
- 21: /cgn2_6/ptodata/1/paa/US097_COMB.pep.*
- 22: /cgn2_6/ptodata/1/paa/US098_COMB.pep.*
- 23: /cgn2_6/ptodata/1/paa/US099_COMB.pep.*
- 24: /cgn2_6/ptodata/1/paa/US100_COMB.pep.*
- 25: /cgn2_6/ptodata/1/paa/US101_COMB.pep.*
- 26: /cgn2_6/ptodata/1/paa/US60_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	545	100.0	100	22	US-09-893-499-3
2	545	100.0	100	22	US-09-893-600-3
3	545	100.0	100	22	US-09-894-030-3
4	304	55.8	56	1	PCT-US98-14547-5
5	304	55.8	56	1	PCT-US98-14547-5
6	304	55.8	56	18	US-09-462-478A-5
7	211	38.7	37	1	PCT-US98-14547-7

8	211	38.7	37	1	PCT-US98-14547-7	Sequence 7, Appli
9	211	38.7	37	18	US-09-462-478A-7	Sequence 7, Appli
10	202	37.1	37	1	PCT-US98-14547-18	Sequence 18, Appli
11	202	37.1	37	1	PCT-US98-14547-18	Sequence 18, Appli
12	202	37.1	37	18	US-09-462-478A-18	Sequence 18, Appli
13	144	26.4	396	4	US-08-085-126-56	Sequence 56, Appli
14	144	26.4	396	8	US-08-438-114-56	Sequence 56, Appli
15	143	26.2	276	18	US-09-417-507-22220	Sequence 22220, A
16	142.5	26.1	918	1	PCT-US99-05250-11	Sequence 11, Appli
17	142.5	26.1	919	19	US-09-515-886-2	Sequence 2, Appli
18	142.5	26.1	919	19	US-09-538-092-895	Sequence 895, App
19	142.5	26.1	923	18	US-09-497-822-19	Sequence 19, Appli
20	142.5	26.1	923	18	US-09-497-822A-19	Sequence 19, Appli
21	141	25.9	175	18	US-09-417-507-43604	Sequence 43604, A
22	141	25.9	209	21	US-09-708-427-10485	Sequence 10485, A
23	141	25.9	221	21	US-09-708-427-10484	Sequence 10484, A
24	140	25.7	94	1	PCT-US00-29360-118	Sequence 118, App
25	140	25.7	94	1	PCT-US01-11988-197	Sequence 197, App
26	140	25.7	94	1	PCT-US02-08123-1836	Sequence 1836, Ap
27	140	25.7	94	1	PCT-US02-08277-1207	Sequence 1207, Ap
28	140	25.7	94	1	PCT-US02-08278-1752	Sequence 1752, Ap
29	140	25.7	94	22	US-09-833-245-197	Sequence 197, App
30	139.5	25.6	285	18	US-09-417-507-22221	Sequence 2221, A
31	138	25.3	188	18	US-09-417-507-22219	Sequence 2219, A
32	135.5	24.9	240	7	US-08-333-576A-30	Sequence 30, Appli
33	135.5	24.9	240	7	US-08-362-670-30	Sequence 30, Appli
34	134.5	24.7	101	1	PCT-US01-00663-26550	Sequence 26550, A
35	134.5	24.7	101	22	US-09-864-761-33599	Sequence 33599, A
36	134.5	24.7	101	26	US-60-236-359-14971	Sequence 14971, A
37	134.5	24.7	1084	17	US-09-394-272-9	Sequence 9, Appli
38	134	24.6	708	16	US-09-248-796-16456	Sequence 16456, A
39	134	24.6	708	26	US-60-096-409-16456	Sequence 16456, A
40	133.5	24.5	169	18	US-09-417-507-25745	Sequence 25745, A
41	133	24.4	252	18	US-09-417-507-43603	Sequence 43603, A
42	133	24.4	262	21	US-09-733-089-9222	Sequence 9222, Ap
43	133	24.4	262	22	US-09-816-660-9222	Sequence 9222, Ap
44	132.5	24.3	201	1	PCT-US01-00358-5	Sequence 5, Appli
45	132.5	24.3	201	1	PCT-US99-17885-11	Sequence 11, Appli

ALIGNMENTS

RESULT 1

US-09-893-499-3
; Sequence 3, Application US/09893499
; GENERAL INFORMATION:
; APPLICANT: Hansen, J. Norman
; TITLE OF INVENTION: Construction and Screening of Lantibody Display Libraries
; FILE REFERENCE: 108172-00056
; CURRENT APPLICATION NUMBER: US/09/893,499
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: 60/215,449
; PRIOR FILING DATE: 2000-06-29
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 100
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: The peptide sequence of sunA-PC20-SL.
US-09-893-499-3

Query Match	100.0%;	Score 545;	DB 22;	Length 100;
Best Local Similarity	100.0%;	Pred. No. 1.4e-49;		
Matches	100;	Conservative	0;	Mismatches
			0;	Indels
				Gaps
				0;
Qy	1	MEKLFKEVKLEENQKSGGLGKRAQCAALWLQCAASGGTICGGGAVACQNYRQFCRGGG	60	
Db	1	MEKLFKEVKLEENQKSGGLGKRAQCAALWLQCAASGGTICGGGAVACQNYRQFCRGGG	60	
Qy	61	GGGGGGGGGGGGGGGSKFDLDFDLVVKVSKQDKITPQ	100	

[illegible]

```

RESULT 2
US-09-893-600-3
; Sequence 3, Application US/09893600
; GENERAL INFORMATION:
; APPLICANT: Hansen, J. Norman
; TITLE OF INVENTION: Construction of a Strain of Bacillus subtilis 168 that Displays
; TITLE OF INVENTION: Sublancin Lantibiotic on the Surface of the Cell
; FILE REFERENCE: 108172-00057
; CURRENT APPLICATION NUMBER: US/09/893,600
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: 60/215,449
; PRIOR FILING DATE: 2000-06-29
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 100
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: The peptide sequence of sunA-PG20-SL.
US-09-893-600-3

Query Match          100.0%; Score 545; DB 22; Length 100;
Best Local Similarity 100.0%; Pred. No. 1.4e-49;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEKLFKEVKLELELNQKSGGLGKAQCAALWLQCAASGTTICGCGGAVACQNYRQFCRGGG 60
   |||||||
Db 1 MEKLFKEVKLELELNQKSGGLGKAQCAALWLQCAASGTTICGCGGAVACQNYRQFCRGGG 60
   |||||||

QY 61 GGGGGGGGGGGGGGSKKDFDLDVVKYSKODSKITPQ 100
   |||||||
Db 61 GGGGGGGGGGGGGGSKKDFDLDVVKYSKODSKITPQ 100
   |||||||

RESULT 3
US-09-894-030-3
; Sequence 3, Application US/09894030
; GENERAL INFORMATION:
; APPLICANT: Hansen, J. Norman
; TITLE OF INVENTION: Construction of a Structural Variant of Sublancin to Facilitate
; TITLE OF INVENTION: Isolation and Use in Bioremediation of Environmental Contaminants
; TITLE OF INVENTION: positive Spore Formers such as Bacillus anthracis
; FILE REFERENCE: 108172-00058
; CURRENT APPLICATION NUMBER: US/09/894,030
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: 60/215,449
; PRIOR FILING DATE: 2000-06-29
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 100
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: The peptide sequence for sublancin-His Tag.
US-09-894-030-3

Query Match          100.0%; Score 545; DB 22; Length 100;
Best Local Similarity 100.0%; Pred. No. 1.4e-49;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEKLFKEVKLELELNQKSGGLGKAQCAALWLQCAASGTTICGCGGAVACQNYRQFCRGGG 60
   |||||||
Db 1 MEKLFKEVKLELELNQKSGGLGKAQCAALWLQCAASGTTICGCGGAVACQNYRQFCRGGG 60
   |||||||

QY 61 GGGGGGGGGGGGGGSKKDFDLDVVKYSKODSKITPQ 100
   |||||||
Db 61 GGGGGGGGGGGGGGSKKDFDLDVVKYSKODSKITPQ 100
   |||||||

```

; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 56
; TYPE: PRT
; ORGANISM: Bacillus subtilis 168
; FEATURE:
; OTHER INFORMATION: Pre-Sublancin 168
US-09-462-478A-5

Query Match 55.8%; Score 304; DB 18; Length 56;
Best Local Similarity 100.0%; Pred. No. 2.2e-24;
Matches 56; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEKLFKEVKEELENKQSGGLGKAQCAALWLCASGGTTCGGGAVACQNYRQFCR 56
|||||
DB 1 MEKLFKEVKEELENKQSGGLGKAQCAALWLCASGGTTCGGGAVACQNYRQFCR 56

RESULT 7
PCT-US98-14547-7
; Sequence 7, Application PC/TUS9814547B
; GENERAL INFORMATION:
; APPLICANT: UNIVERSITY OF MARYLAND
; TITLE OF INVENTION: SUBLANCIN LANTIBIOTIC PRODUCED BY BACILLUS SUBTILIS 168
; FILE REFERENCE: 8172-8068
; CURRENT APPLICATION NUMBER: PCT/US98/14547B
; CURRENT FILING DATE: 1998-07-17
; EARLIER APPLICATION NUMBER: US/60/053,035
; EARLIER FILING DATE: 1997-07-18
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 37
; TYPE: PRT
; ORGANISM: Bacillus subtilis 168
; FEATURE:
; OTHER INFORMATION: Prosublancin 168
PCT-US98-14547-7

Query Match 38.7%; Score 211; DB 1; Length 37;
Best Local Similarity 100.0%; Pred. No. 9.3e-15;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 GLGKAQCAALWLCASGGTTCGGGAVACQNYRQFCR 56
|||||
DB 1 GLGKAQCAALWLCASGGTTCGGGAVACQNYRQFCR 37

RESULT 8
PCT-US98-14547-7
; Sequence 7, Application PC/TUS9814547C
; GENERAL INFORMATION:
; APPLICANT: UNIVERSITY OF MARYLAND
; TITLE OF INVENTION: SUBLANCIN LANTIBIOTIC PRODUCED BY BACILLUS SUBTILIS 168
; FILE REFERENCE: 8172-8072
; CURRENT APPLICATION NUMBER: PCT/US98/14547C
; CURRENT FILING DATE: 1998-07-17
; EARLIER APPLICATION NUMBER: US/60/053,035
; EARLIER FILING DATE: 1997-07-18
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 37
; TYPE: PRT
; ORGANISM: Bacillus subtilis 168
; FEATURE:
; OTHER INFORMATION: Prosublancin 168
PCT-US98-14547-7

Query Match 38.7%; Score 211; DB 1; Length 37;
Best Local Similarity 100.0%; Pred. No. 9.3e-15;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 GLGKAQCAALWLCASGGTTCGGGAVACQNYRQFCR 56
|||||
DB 1 GLGKAQCAALWLCASGGTTCGGGAVACQNYRQFCR 37

RESULT 9
US-09-462-478A-7
; Sequence 7, Application US/09462478A
; GENERAL INFORMATION:
; APPLICANT: UNIVERSITY OF MARYLAND
; TITLE OF INVENTION: SUBLANCIN LANTIBIOTIC PRODUCED BY BACILLUS SUBTILIS 168
; FILE REFERENCE: 8172-8072
; CURRENT APPLICATION NUMBER: US/09/462,478A
; CURRENT FILING DATE: 2000-04-17
; PRIOR APPLICATION NUMBER: US/60/053,035
; PRIOR FILING DATE: 1997-07-18
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 37
; TYPE: PRT
; ORGANISM: Bacillus subtilis 168
; FEATURE:
; OTHER INFORMATION: Prosublancin 168
US-09-462-478A-7

Query Match 38.7%; Score 211; DB 18; Length 37;
Best Local Similarity 100.0%; Pred. No. 9.3e-15;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 GLGKAQCAALWLCASGGTTCGGGAVACQNYRQFCR 56
|||||
DB 1 GLGKAQCAALWLCASGGTTCGGGAVACQNYRQFCR 37

RESULT 10
PCT-US98-14547-18
; Sequence 18, Application PC/TUS9814547B
; GENERAL INFORMATION:
; APPLICANT: UNIVERSITY OF MARYLAND
; TITLE OF INVENTION: SUBLANCIN LANTIBIOTIC PRODUCED BY BACILLUS SUBTILIS 168
; FILE REFERENCE: 8172-8068
; CURRENT APPLICATION NUMBER: PCT/US98/14547B
; CURRENT FILING DATE: 1998-07-17
; EARLIER APPLICATION NUMBER: US/60/053,035
; EARLIER FILING DATE: 1997-07-18
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 37
; TYPE: PRT
; ORGANISM: Bacillus subtilis 168
; FEATURE:
; NAME/KEY: DISULFID
; LOCATION: (14)..(29)
; FEATURE:
; NAME/KEY: DISULFID
; LOCATION: (7)..(36)
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (16)
; OTHER INFORMATION: Xaa is a dehydrogenated Ser (Dha)
; FEATURE:
; NAME/KEY: THIOETH
; LOCATION: (19)..(22)
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (19)
; OTHER INFORMATION: Xaa is a dehydrogenated Thr (Dhb)
PCT-US98-14547-18

Query Match 37.1%; Score 202; DB 1; Length 37;

```

RESULT 12
US-09-462-478A-18
; Sequence 18, Application US/09462478A
; GENERAL INFORMATION:
; APPLICANT: UNIVERSITY OF MARYLAND
; TITLE OF INVENTION: SUBLANCIN LANTIBIOTIC PRODUCED BY BACILLUS SUBTILIS 168
; FILE REFERENCE: 8172-8072
; CURRENT APPLICATION NUMBER: US/09/462,478A
; CURRENT FILING DATE: 2000-04-17
; PRIOR APPLICATION NUMBER: US/60/053,035
; PRIOR FILING DATE: 1997-07-18
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 37
; TYPE: PRT
; ORGANISM: Bacillus subtilis 168
; FEATURE:
; NAME/KEY: DISULFID
; LOCATION: (14)..(29)

```

```

Query Match      26.4%; Score 144; DB 4; Length 396;
Best Local Similarity 53.1%; Pred. No. 1.6e-06;
Matches 26; Conservative 6; Mismatches 17; Indels

QY 36 GTTIGCGGAVACQNYRFCRGCGGGGGGGGGGGGGGKFFDFD 84
    || || : : | | | | | | | | | | | | | |
Db 104 GCGMGNNGNFKAEIDMRTCTGAGGGGGGGGGGGGGGAAPSF 152

```

RESULT 14
US-08-438-114-56

RESULT 15
US-09-417-507-22220
; Sequence 22220, Application US/09417507
; GENERAL INFORMATION:
; APPLICANT: KEITH G. WEINSTOCK ET AL.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ASPERGILLUS
; TITLE OF INVENTION: FUMIGATUS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: PATH99-10
; CURRENT APPLICATION NUMBER: US/09/417,507
; CURRENT FILING DATE: 1999-10-14
; NUMBER OF SEQ ID NOS: 44312
; SEQ ID NO 22220
; LENGTH: 276
; TYPE: PRT
; ORGANISM: A.fumigatus
US-09-417-507-22220

```

Best Local Similarity 40.7%; Pred. No. 1.4e-06;
Matches 33; Conservative 3; Mismatches 23; Indels 22; Gaps 3;

Qy 20 GLAKAOCALWLQCA----SGGTICGGGAVACONRYQRCGGGGGGGGGG-----69
      | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 185 GMAISACLCAWFDSABGVGEGGGRGGGGGRSGWG-----CGGGGGGGGAGGGAWGEGG 240

Qy 70 -----GGGGGGGMSKDD 82
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 241 RGGGGWGRGGGGGGGCGQCFDD 261

Search completed: October 24, 2002, 09:33:47.
Job time : 154.555 secs

```


Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	142.5	26.1	918	4	US-09-041-886-11	Sequence 11, Appl
2	135.5	24.9	240	1	US-08-362-670B-30	Sequence 30, Appl
3	135.5	24.9	240	3	US-08-333-576C-30	Sequence 30, Appl
4	135.5	24.9	240	4	US-08-808-32C-30	Sequence 30, Appl
5	135.5	24.9	240	5	PTC-US94-14030A-30	Sequence 30, Appl
6	132.5	24.3	201	4	US-09-052-995-1	Sequence 1, Appl
7	132.5	24.3	201	4	US-09-053-003-40	Sequence 40, Appl
8	129.5	23.8	738	3	US-08-864-038A-3	Sequence 3, Appl
9	129	23.7	268	2	US-08-835-039A-9	Sequence 9, Appl
10	129	23.7	268	3	US-09-157-349-9	Sequence 9, Appl
11	125	22.9	161	2	US-08-581-528A-6	Sequence 6, Appl
12	125	22.9	161	5	PTC-US94-07799-6	Sequence 6, Appl
13	125	22.9	266	4	US-09-032-523-7	Sequence 7, Appl
14	124	22.8	322	1	US-08-014-943A-2	Sequence 2, Appl
15	124	22.8	322	1	US-08-486-421-3	Sequence 3, Appl
16	124	22.8	322	1	US-08-470-911-3	Sequence 3, Appl
17	124	22.8	322	2	US-08-486-809-3	Sequence 3, Appl
18	123.5	22.7	148	1	US-08-207-904-15	Sequence 15, Appl
19	122.5	22.5	400	4	US-09-086-010-2	Sequence 2, Appl
20	121	22.2	24	1	US-08-524-677-11	Sequence 11, Appl
21	121	22.2	24	2	US-08-465-491-26	Sequence 26, Appl
22	121	22.2	24	2	US-08-986-617-26	Sequence 26, Appl
23	121	22.2	56	1	US-07-981-525-2	Sequence 2, Appl
24	121	22.2	56	1	US-07-981-525-7	Sequence 7, Appl
25	121	22.2	56	1	US-07-981-525-9	Sequence 9, Appl
26	121	22.2	56	1	US-08-220-033-2	Sequence 2, Appl
27	121	22.2	56	1	US-08-220-033-7	Sequence 7, Appl

NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:

SIRHAN, 67 CAMDENWAY PARK DRIVE
CITY: Cambridge
STATE: Massachusetts
COUNTRY: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/333,576C
FILING DATE: NO. 6027919ember 2, 1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:

NAME: Lazar, Steven R.
REGISTRATION NUMBER: 32,618
REFERENCE/DOCKET NUMBER: 5202-B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617 498-8260
TELEFAX: 617 876-5851
INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:
LENGTH: 240 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
-333-576C-30

```

ry Match      24.9%; Score 135.5; DB 3; Length 240;
t Local Similarity 40.4%; Pred.No.7.2e-07;
ches 36; Conservative 5; Mismatches 25; Indels 23
2 EKLFKEVKLE-----ELEKNGSGLG--KAQCAALWLO-CASGGTIGCGGGAUVA
| ||||: : | ||||| | ||||| | |||||
53 ESUFREIRAQRALRAAAPPPDPFGAGSRKANLGGRRRORTALACTRGXXGS
| ||||| | ||||| | ||||| | ||||| | |||||
51 YROFCRGGGGGGGGGGGGGGGGGGMSK 79
| ||||| | ||||| | ||||| | ||||| | |||||
107 -----GGGGGGGGGGGGGGGGGAGR 129

```

GENERAL INFORMATION:
 APPLICANT: Celeste, Anthony J.
 License No. 6284872
 License Issued: 05/08/000524

APPLICANT: Rosen, Vicki A.
 APPLICANT: Wolfman, Neil
 APPLICANT: Thomsen, Gerald H.
 APPLICANT: Melton, Douglas A.
 TITLE OF INVENTION: TENDON-INDUCING COMPOSITIONS
 NUMBER OF SEQUENCES: 35
 CORRESPONDENCE ADDRESS:

STREET: 67 Cambridgepark Drive
 CITY: Cambridge
 STATE: Massachusetts
 COUNTRY: USA
 ZIP: 02140
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/808,324
 FILING DATE: Herewith

APPLICATION NUMBER: US/08/808,324
FILING DATE: Herewith

```

PCT-US94-14030A-30

Query Match          24.9%  Score 135.5;  DB 5;  Length 240;
Best Local Similarity 40.4%  Pred. No. 7.2e-07;
Matches 36;  Conservative 5;  Mismatches 25;  Indels 23;  Gaps 4;

QY      2  EKLFEVKLE-----ELENQKSGSLG--KAQCAALWLO-CASGGTTIGCGGAVACQN 50
      | | | | | : | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      53  ESLFREIRQAARALPAARAEPPDPGPGAGSRKANILGRRRQRTALAGTFRXXGS----- 106
      | | | | | : | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY      51  YRQFCRGGGGGGGGGGGGGGGGGGGMSK 79
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      107  -----GGGGGGGGGGGGGGGGGGGAGR 129
      | | | | | : | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 6
US-09-052-995-1
; Sequence 1, Application US/09052995
; Patent No. 6183956
; GENERAL INFORMATION:
; APPLICANT: Sivarama, Mohanram
; APPLICANT: Strulovici, Berta
; APPLICANT: Flores, Osvaldo A.
; TITLE OF INVENTION: High Throughput In Vitro Screening Assay
; TITLE OF INVENTION: for Transcription Modulators
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/052,995
; FILING DATE: 31-MAR-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Parent, Annette S.
; REGISTRATION NUMBER: 42,058
; REFERENCE/DOCKET NUMBER: 018781-00060005
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 201 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-052-995-1

Query Match          24.3%  Score 132.5;  DB 4;  Length 201;
Best Local Similarity 63.4%  Pred. No. 1.2e-06;
Matches 26;  Conservative 0;  Mismatches 4;  Indels 11;  Gaps 1;

QY      36  GTTICGGGAVACQNYRQFCRGGGGGGGGGGGGGGGGGG 76
      | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      99  GPGGGGGGGG-----GGGGGGGGGGGGGGGGGG 128
      | | | | | : | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 7
US-09-053-003-40
; Sequence 40, Application US/09053003
; Patent No. 6207391
; GENERAL INFORMATION:
; APPLICANT: Wu, Penquang

```

```

; APPLICANT: McKinney, Judi
; TITLE OF INVENTION: High-Throughput Screening Assays for
; TITLE OF INVENTION: Modulators of STAT4 and STAT6 Activity
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/053,003
; FILING DATE: 31-MAR-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, Timothy L.
; REGISTRATION NUMBER: 35,367
; REFERENCE/DOCKET NUMBER: 018781-000800US
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 201 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 1..97
; OTHER INFORMATION: /product= "OTHER"
; OTHER INFORMATION: /note= "Gly at positions 1-97 may be
; OTHER INFORMATION: present or absent"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 105..201
; OTHER INFORMATION: /product= "OTHER"
; OTHER INFORMATION: /note= "Gly at positions 105-201 may be
; OTHER INFORMATION: present or absent"
;
; US-09-053-003-40
;
; Query Match 24.3%; Score 132.5; DB 4; Length 201;
; Best Local Similarity 63.4%; Pred. No. 1.2e-06;
; Matches 26; Conservative 0; Mismatches 4; Indels 11; Gaps 1;
;
; QY 36 GGTGCGGAVACQNYRQFCRGGGGGGGGGGGGGGGGGGGG 76
; Db 99 GCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 128
;
; RESULT 8
; US-08-864-038A-3
; Sequence 3, Application US/08864038A
; Patent No. 6001592
; GENERAL INFORMATION:
; APPLICANT: Kunio NAKASHIMA et al.
; TITLE OF INVENTION: NOVEL POLYPEPTIDE GENE CDNA, VECTOR
; TITLE OF INVENTION: CONTAINING SAID CDNA, HOST CELLS TRANSFORMED WITH SAID
; TITLE OF INVENTION: VECTOR, POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING
; TITLE OF INVENTION: SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE AND ANTIBODY
; TITLE OF INVENTION: TO SAID POLYPEPTIDE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: 812-5 Hirano
; STREET: Isshinden
; CITY: Tsu-city

```

```

; STATE: Mie-prefecture
; COUNTRY: JAPAN
; ZIP: 514-01
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Microsoft Windows 95
; SOFTWARE: Word Perfect 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/864,038A
; FILING DATE: May 28, 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-184459
; FILING DATE: 15-July-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: C. Bruce Hamburg
; REGISTRATION NUMBER: 22,389
; REFERENCE/DOCKET NUMBER: F-5610
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)986-2340
; TELEFAX: (212)953-7733
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 738
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Pinctada fucata
; CELL TYPE: mantle epithelial cell
; FEATURE:
; NAME/KEY: peptide
; LOCATION: from 1 to 738
; IDENTIFICATION METHOD: E (by experiment)
;
; US-08-864-038A-3
;
; Query Match 23.8%; Score 129.5; DB 3; Length 738;
; Best Local Similarity 41.0%; Pred. No. 9.2e-06;
; Matches 32; Conservative 3; Mismatches 12; Indels 31; Gaps 3;
;
; QY 18 GSGGLKACCAALWLOCASGGTIGCGGAVACQNYRQFCRGGGGGGGGGGGGGGGGGGGG-- 75
; Db 674 GSGGG-----GSGGGGGGGGG-----GSGGGGGGGGGGGGGGGGGGG 711
;
; QY 76 -----GMSKFDFFDLD 86
; Db 712 NNGWNGNNGNYDDDDCD 729
;
; RESULT 9
; US-08-835-099A-9
; Sequence 9, Application US/08835099A
; Patent No. 5874277
; GENERAL INFORMATION:
; APPLICANT: SHINTANI, Yasushi
; APPLICANT: NISHI, Kazuori
; APPLICANT: KAWAMOTO, Tomohiro
; TITLE OF INVENTION: NOVEL PROTEINS, THEIR PRODUCTION
; TITLE OF INVENTION: AND USE
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSMAN, LLP
; STREET: 130 Water Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:

```

```

RESULT 10
US-09-157-349-9
/ Sequence 9, Application US/09157349
/ Patent No. 6068990
/ GENERAL INFORMATION:
/ APPLICANT: SHINTANI, Yasushi
/ APPLICANT: SHINTAI, Kazunori
/ APPLICANT: KAWAMOTO, Tomohiro
/ TITLE OF INVENTION: NOVEL PROTE
/ TITLE OF INVENTION: AND USE
/ NUMBER OF SEQUENCES: 18
/ CORRESPONDENCE ADDRESS:
/ ADDRESSER: DIKE, BRONSTEIN, RO
/ STREET: 130 Water Street
/ CITY: Boston
/ STATE: MA
/ COUNTRY: USA
/ ZIP: 02109
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Diskette
/ COMPUTER: IBM Compatible
/ OPERATING SYSTEM: DOS
/ SOFTWARE: FastSEQ for Windows
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/157
/ FILING DATE:
/ CLASSIFICATION:
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 08/835,095
/ FILING DATE:
/ APPLICATION NUMBER: 97105508.1
/ FILING DATE: 03-APR-1997
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Resnick, David S
/ REGISTRATION NUMBER: 34,235
/ REFERENCE/DOCKET NUMBER: 4734
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 617-523-3400
/ TELEFAX: 617-523-6440

```

RESULT 12

PCT-US94-07799-6
; Sequence 6, Application PC/TUS9407799
; GENERAL INFORMATION:
; APPLICANT: THE JOHNS HOPKINS UNIVERSITY SCHOOL OF MEDICINE
; TITLE OF INVENTION: GROWTH DIFFERENTIATION FACTOR-7
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Spensley Horn Juba & Lubitz
; STREET: 1880 Century Park East, Suite 500
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90067
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/07799
; FILING DATE: 08-JUL-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: TUMARKIN, LISA A., PH.D.
; REGISTRATION NUMBER: P-38,347
; REFERENCE/DOCKET NUMBER: FD-2348
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 455-5100
; TELEFAX: (619) 455-5110
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 161 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US94-07799-6
Query Match 22.9%; Score 125; DB 5; Length 161;
Best Local Similarity 54.3%; Pred. No. 5.8e-06;
Matches 25; Conservative 1; Mismatches 8; Indels 12; Gaps 1;
QY 34 ASGTGCGGGAACACNYRQFCRGGGGGGGGGGGGGGGMSK 79
DB 17 ALAGTRGAQGS-----GGGGGGGGGGGGGGGGGGAGR 50
RESULT 13
US-09-032-523-7
; Sequence 7, Application US/09032523
; Patent No. 6232454
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Corley, Neil C.
; APPLICANT: Guegler, Karl
; APPLICANT: Baugh, Mariah
; TITLE OF INVENTION: HUMAN PROTEINASE MOLECULES
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/032,523
; FILING DATE:

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0479 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 266 amino acids
TYPE: amino acid
STRAINEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 164403
US-09-032-523-7
Query Match 22.9%; Score 125; DB 4; Length 266;
Best Local Similarity 44.6%; Pred. No. 9.6e-06;
Matches 29; Conservative 3; Mismatches 9; Indels 24; Gaps 3;
QY 17 KSGLGKACQCAALWLCQASGGTIGG-----GGAVACNYRQFCRGGGGGGGGGGGGG 71
DB 9 KGGGGGG-----GGGGGLGGLNVLGLIS-----GAGGGGGGGGGGGGGG 49
QY 72 GGGGG 76
DB 50 GGGGG 54
RESULT 14
US-08-014-943A-2
; Sequence 2, Application US/08014943A
; Patent No. 5545551
; GENERAL INFORMATION:
; APPLICANT: Johnson, Edward M.
; APPLICANT: Bergemann, Andrew D.
; TITLE OF INVENTION: Cloning And Expression Of Pur Protein
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/014,943A
; FILING DATE: 02/FEB/1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 6923-033
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 322 amino acids
; TYPE: amino acid

Search completed: October 24, 2002, 09:30:12
Job time : 18.3285 secs

```

; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-014-943A-2

Query Match      22.88; Score 124; DB 1; Length 322;
Best Local Similarity 38.08; Pred. NO. 1.5e-05;
Matches 38; Conservative 6; Mismatches 10; Indels 46; Gaps 6;

QY 16 QKGSGLGKRAQCAALWLOACSGGTI---CGGGGAVACQNYRQFCRCGGGGGGGGGGGGGG 72
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 10 QGGAALG-----SGGSLGHPGSGGS-----GGGGGGGGGGGGGGGGGGGG 46

QY 73 GGGG-----MSKFD-----FDLDVVKVSKQDSK 96
   ||| || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 47 GGGAPGGLQHETQELASKRVDIQNKRFYLDV---KONAK 82

RESULT 15
US-08-486-421-3
; Sequence 3, Application US/08486421
; Patent No. 5672479
; GENERAL INFORMATION:
; APPLICANT: Johnson, Edward M.
; APPLICANT: Bergemann, Andrew D.
; TITLE OF INVENTION: CLONING AND EXPRESSION OF PUR PROTEIN
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/486,421
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/470,911
; FILING DATE: 06-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 6923-053
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 322 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-486-421-3

Query Match      22.88; Score 124; DB 1; Length 322;
Best Local Similarity 38.08; Pred. NO. 1.5e-05;
Matches 38; Conservative 6; Mismatches 10; Indels 46; Gaps 6;

QY 16 QKGSGLGKRAQCAALWLOACSGGTI---CGGGGAVACQNYRQFCRCGGGGGGGGGGGGGG 72
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 10 QGGAALG-----SGGSLGHPGSGGS-----GGGGGGGGGGGGGGGGGGGG 46

QY 73 GGGG-----MSKFD-----FDLDVVKVSKQDSK 96
   ||| || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 47 GGGAPGGLQHETQELASKRVDIQNKRFYLDV---KONAK 82
```


Thu Oct 24 14:40:01 2002

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model
Run on: October 24, 2002, 09:26:16 ; Search time 18.9781 Seconds
(without alignments)
506.317 Million cell updates/sec

Title: US-09-894-030-3
Perfect score: 545
Sequence: 1 MEKLFKEVKELEENQKSG.....DDFDLVVVKQDSKITPQ 100

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 71:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	304	55.8	56	T12783	sublancin 168 prec
2	165	30.3	336	S18750	chitinase (EC 3.2.
3	151	27.7	385	T20410	hypothetical prote
4	142.5	26.1	919	A39248	androgen receptor
5	141	25.9	221	T04592	glycine-rich cell
6	139.5	25.6	393	T20268	hypothetical prote
7	138	25.3	207	T07381	glycine-rich prote
8	134.5	24.7	1084	T04103	sucrose-phosphate
9	132.5	24.3	433	S20963	homeotic protein H
10	130	23.9	165	KNRZG1	glycine-rich cell
11	130	23.9	255	B84777	hypothetical prote
12	130	23.9	291	S31415	glycine-rich prote
13	129	23.7	106	F84797	hypothetical prote
14	129	23.7	268	CIHUL	calpain (EC 3.4.22
15	129	23.7	431	WJHU2G	homeotic protein H
16	128	23.5	214	KNNTZS	glycine-rich prote
17	128	23.5	681	AB2155	hypothetical prote
18	127	23.3	271	S34666	glycine-rich prote
19	127	23.3	384	A26099	glycine-rich cell
20	127	23.3	396	T49109	hypothetical prote
21	126	23.1	290	AD1849	homeotic protein E
22	126	23.1	333	A39085	RNA-binding protei
23	125.5	23.0	404	S54729	endo-1,4-beta-gluc
24	125.5	23.0	592	B82759	bone morphogenetic
25	125	22.9	151	S43296	eggshell protein p
26	125	22.9	220	A44805	calpain (EC 3.4.22
27	125	22.9	266	CIPLGL	fibrillarlin [valid
28	124.5	22.8	321	A38712	glycine-rich prote
29	124	22.8	201	F84596	

30	124	22.8	322	2	A45036	single-stranded-DN
31	123.5	22.7	152	2	G96010	hypothetical expor
32	123	22.6	280	2	A42424	chitinase (EC 3.2.
33	123	22.6	1585	2	T31611	hypothetical prote
34	122.5	22.5	336	2	T29282	hypothetical prote
35	122	22.4	388	2	T29173	hypothetical prote
36	121.5	22.3	171	2	H84709	probable glycine-r
37	121.5	22.3	268	2	S09860	hypothetical prote
38	121.5	22.3	302	2	C84470	hypothetical prote
39	121.5	22.3	316	1	A38743	loricrin - human
40	121.5	22.3	481	2	A35628	loricrin - mouse
41	121.5	22.3	549	2	B86264	hypothetical prote
42	121	22.2	56	1	NIBSSA	subtilin precursor
43	120.5	22.1	199	2	S16063	acp-22 protein - y
44	120.5	22.1	199	2	S32224	acp-22 protein - y
45	120.5	22.1	266	1	CIRBL	calpain (EC 3.4.22

ALIGNMENTS

RESULT 1

T12783
sublancin 168 precursor - Bacillus subtilis phage SPBc2
C:Species: Bacillus subtilis phage SPBc2
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 24-Sep-1999
C:Accession: T12783; H69719
R:Lazarevic, V.; Duesterhoeft, A.; Solido, B.; Hilbert, H.; Mauel, C.; Karamata, D.
submitted to the EMBL Data Library, August 1997
A:Description: The complete nucleotide sequence of the Bacillus subtilis SPbetac2 pro
A:Reference number: Z17583
A:Accession: T12783
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-56 <LAZ>
A:Cross-references: EMBL:AF020713; NID:g3025478; PID:g3025497; PIDN:AAC12992.1
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
C.; Bron, S.; Brouillet, S.; Bruchli, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, M.; Rose, M.; Sadale, Y.; Sato, T.; Scanl
A:Authors: Schleich, S.; Schroeter, R.; Scroffone, F.; Sekiguchi, J.; Sekowska, A.; Se
keuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Togononi, A.; Tosato, V.; Uchiya
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtili
A:Reference number: A69580; MUID:98044033
A:Accession: H69719
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-56 <KUN>
A:Cross-references: GB:Z99115; GB:AL009126; NID:g2634478; PIDN:CAB14066.1; PID:el1835
A:Experimental source: strain 168
C:Genetics: <LAI>
A:Gene: yolg
C:Genetics: <KUI>
A:Gene: sunA
C:Superfamily: unassigned lanthionine-containing peptides
C:Keywords: antibiotic; lanthionine
F;1-19/Domain: propeptide #status predicted <PRO>
F;20-56/Product: sublancin 168 #status predicted <MAT>
F;26-55/Disulfide bonds: #status experimental
F;33-48/Disulfide bonds: #status predicted
F;35/Modified site: dehydroalanine (Ser) #status experimental
F;38-41/Cross-link: (2S,3S,6R)-3-methyl-lanthionine (Thr-Cys) #status predicted

Query Match 55.8%; Score 304; DB 2; Length 56;
Best Local Similarity 100.0%; Pred. No. 4.4e-22;

RESULT 2
S18750
chitinase (EC 3.2.1.14) precursor - western balsam poplar x cottonwood
C:Species: Populus trichocarpa x Populus deltoides (western balsam poplar x cottonwood)
C:date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: S18750; S17755

A;Residues: 1-919 <D3>
A;Cross-references: GB:M20132; PIDN:AAA51729.1; PID:gl78628; GB:J03180
R;Trapman, J.; Klaassen, P.; Kuiper, G.G.J.M.; van der Korput, J.A.G.M.; Faber, P.W.;
Biochem. Biophys. Res. Commun. 153, 241-248, 1988
A;Title: Cloning, structure and expression of a cDNA encoding the human androgen recep
A;Reference number: A27653; MUID:88240407
A;Accession: A27653
A;Molecule type: mRNA
A;Residues: 468-564, 'K', 566-919 <TRA>
A;Cross-references: GB:M20260; PID:gl78891; PIDN:AAA51774.1; PID:gl78892
A;Note: the authors translated the codon AAG for residue 565 as Glu
R;Chang, C.; Kokontis, J.; Liao, S.
Science 240, 324-326, 1988
A;Title: Molecular cloning of human and rat complementary DNA encoding androgen recep
A;Reference number: A40108; MUID:88178111
A;Accession: A40108
A;Molecule type: mRNA
A;Residues: 557-628 <CHA>
A;Cross-references: GB:M18624
R;Chang, C.; Kokontis, J.; Liao, S.
Proc. Natl. Acad. Sci. U.S.A. 85, 7211-7215, 1988

A;Title: Structural analysis of complementary DNA and amino acid sequences of human and

A;Reference number: A40494; MUID:89017108

A;Accession: A40494

A;Molecule type: mRNA

A;Residues: 1-74, 79-89, 'H', 90-472, 'GGG', 473-474, 'E', 476-644, 'N', 646-919 <CH2>

A;Cross-references: GB:M23263

R;Tilley, W.D.; Marcelli, M.; Willson, J.D.; McPhaul, M.J.

Proc. Natl. Acad. Sci. U.S.A. 86, 327-331, 1989

A;Title: Characterization and expression of a cDNA encoding the human androgen receptor.

A;Reference number: A32224; MUID:89098909

A;Accession: A32224

A;Molecule type: mRNA

A;Residues: 1-77, 79-211, 'R', 213-471, 473-919 <TTL>

A;Cross-references: GB:M21748; GB:J04150; PID:q178871; PIDN:AAA51771.1; PID:q178872

R;Kowicz, I.; Lee, H.J.; Chen, H.T.; Mestayer, C.; Portois, M.C.; Cabrol, S.; Mauvais

Mol. Endocrinol. 7, 861-869, 1993

A;Title: A point mutation in the second zinc finger of the DNA-binding domain of the and

stance.

A;Reference number: A40715; MUID:94019395

A;Accession: A40715

A;Status: not compared with conceptual translation

A;Molecule type: DNA

A;Residues: 557-614, 'H', 616-624 <MOW>

A;Cross-references: PIDN:AAB28340.1; PID:g4255580

C;Genetics:

A;Gene: GDB:AR

A;Cross-references: GDB:120556; OMIM:313700

A;Map position: Xq11-Xq12

A;Introns: 538/2; 589/2; 628/1; 724/1; 772/2; 816/1; 868/3

C;Superfamily: unassigned erba-related proteins; erba transforming protein homology

C;Keywords: DNA binding; Steroid binding; transcription regulation; zinc finger

F;557-615/Domain: erba transforming protein homology <ERBA>

F;559-579/Region: zinc finger

F;595-619/Region: zinc finger

Query Match 26.1%; Score 142.5; DB 2; Length 919;

Best Local Similarity 42.5%; Pred. No. 7.3e-06;

Matches 37; Conservative 3; Mismatches 22; Indels 25; Gaps 4;

QY 12 ELENKSGSLGKAQCAALWLQACSG-----GTICGGGG---AVACQNYRQF----- 54

DB 386 KLENPLDYGSNWAAAAA--QCRYGLASLHGAGAGPGSGSPSAAASSWHILFTAEAG 442

QY 55 -----CRGGGGGGGGGGGGGGGGGG 76

DB 443 QLYGPCGGGGGGGGGGGGGGGGGG 469

RESULT 5

T04592

glycine-rich cell wall structural protein homolog F23E13.120 - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 14-May-1999

C;Accession: T04592

R;Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Jesse, T.

submitted to the Protein Sequence Database, March 1998

A;Reference number: Z15378

A;Accession: T04592

A;Molecule type: DNA

A;Residues: 1-221 <BEV>

A;Cross-references: EMBL:AL022141

A;Experimental source: cultivar Columbia; BAC clone F23E13

C;Genetics:

A;Map position: 4

A;Note: F23E13.120

Query Match 25.9%; Score 141; DB 2; Length 221;

Best Local Similarity 65.1%; Pred. No. 3.1e-06;

Matches 28; Conservative 1; Mismatches 14; Indels 0; Gaps 0;

QY 34 ASGTTCGGGAVACQNYRQFCRGGGGGGGGGGGGGGGGGG 76

DB 130 SGGGGGGGGGGGSGNGSGRGGGGGGGGGGGGGGGGGGGG 172

A:Residues: 1-1084 <SAK>
A:Cross-references: EMBL:D45890; PIDN:BA08304.1
A:Experimental source: subsp. Japonica
C:Genetics:
A:Gene: Sps1
A:Map position: 1
A:Interons: 120/3; 200/2; 221/3; 452/3; 536/3; 595/3; 613/3; 634/3; 946/3; 989/2
C:Superfamily: sucrose-phosphate synthase; sucrose-phosphate synthase homology
C:Keywords: glycosyltransferase; hexosyltransferase
F:196-680/Domain: sucrose/sucrose-phosphate synthase homology <SSPS>
Query Match 24.7%; Score 134.5; DB 2; Length 1084;
Best Local Similarity 60.5%; Pred. No. 4.8e-05;
Matches 26; Conservative 1; Mismatches 3; Indels 13; Gaps 1;
QY 35 SGGTICGGGAVACQNYRQFCRGGGGGGGGGGGGGGGGGG 77
DB 17 SGAAGGGG-----GGGGGGGGGGGGGGGGGGGG 46
RESULT 9
S20963
homeotic protein Hox B3 - mouse
N:Alternate names: homeotic protein Hox 2.7
C:Species: Mus musculus (house mouse)
C:Date: 22-Nov-1993 #sequence_revision 21-Jul-1995 #text_change 20-Aug-1999
C:Accession: S20963; D42694
R:Sham, M.H.; Hunt, P.; Nonchev, S.; Papalopulu, N.; Graham, A.; Boncinelli, E.; Krumlauf
EMBO J. 11, 1825-1836, 1992
A:Title: Analysis of the murine Hox-2.7 gene: conserved alternative transcripts with dif
A:Reference number: S20963; MUID:92258392
A:Accession: S20963
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-433 <SHA>
A:Cross-references: GB:X66177; GB:S35628; GB:S35738; NID:g312229; PIDN:CAA46951.1; PID:9
R:Nazarali, A.; Kim, Y.; Nirenberg, M.
Proc. Natl. Acad. Sci. U.S.A. 89, 2883-2887, 1992
A:Title: Hox-1.11 and Hox-4.9 homeobox genes.
A:Reference number: A42694; MUID:92212934
A:Accession: D42694
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 213-238 <NAZ>
A:Note: sequence extracted from NCBI backbone (NCBIN:92310, NCBI:P:92316)
C:Superfamily: homeotic protein Hox B3; homeobox homology
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
F:192-248/Domain: homeobox homology <HOX>
Query Match 24.3%; Score 132.5; DB 2; Length 433;
Best Local Similarity 56.6%; Pred. No. 3.4e-05;
Matches 30; Conservative 2; Mismatches 6; Indels 15; Gaps 3;
QY 31 LQCASGGTT-CCGGGAVACQNYRQFCRGGGGGGGGGGGG 79
DB 142 LKNSPGTAEGCGGGG-----GGGGGGGGGGGGGGGGGG 183
RESULT 10
KNRZG1
glycine-rich cell wall structural protein 1 precursor (clone lambda-313) - rice
C:Species: Oryza sativa (rice)
C:Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 16-Jul-1999
C:Accession: S13385
R:Lei, M.; Wu, R.
Plant Mol. Biol. 16, 187-198, 1991
A:Title: A novel glycine-rich cell wall protein gene in rice.
A:Reference number: S13385; MUID:91370862
A:Accession: S13385
A:Molecule type: DNA
A:Residues: 1-165 <LEI>
A:Cross-references: EMBL:X53596; NID:g20246; PIDN:CAA37665.1; PID:g20247
C:Genetics:
A:Gene: grp-1
C:Superfamily: glycine-rich cell wall structural protein 1
C:Keywords: cell wall; duplication; structural protein
F:1-23/Domain: signal sequence #status predicted <SIG>
F:124-165/Product: glycine-rich cell wall structural protein 1 #status predicted <MAT>
F:30-55/Region: repeat R1
F:56-62/Region: repeat R2
F:62-92/Region: repeat R1
F:93-99/Region: repeat R2
F:100-131/Region: repeat R1
F:132-138/Region: repeat R2
F:139-160/Region: repeat R1
Query Match 23.9%; Score 130; DB 1; Length 165;
Best Local Similarity 48.4%; Pred. No. 2.7e-05;
Matches 31; Conservative 3; Mismatches 22; Indels 8; Gaps 2;
QY 18 GSGLGKACAAALWLQCAASGGTIGCGGAVACQNYRQ----FCRGGGGGGGGGGGGGG 72
DB 99 CGNGGAQCGQ---SGGGGGGGGGGGGGGGGGGGGG 155
QY 73 GGGG 76
DB 156 GGS 159
RESULT 11
B84777
hypochemical protein At2g36120 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C:Accession: B84777
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
euss. D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487
A:Accession: B84777
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-255 <STO>
A:Cross-references: GB:AE002093; NID:g4678224; PIDN:AAD26969.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g36120
A:Map position: 2
C:Superfamily: collagen alpha 1(V) chain; fibrillar collagen carboxyl-terminal homolo
Query Match 23.9%; Score 130; DB 2; Length 255;
Best Local Similarity 50.8%; Pred. No. 3.8e-05;
Matches 30; Conservative 0; Mismatches 19; Indels 10; Gaps 1;
QY 18 GSGLGKACAAALWLQCAASGGTIGCGGAVACQNYRQFCRGGGGGGGGGGGGGGGGGG 76
DB 182 GSGEGGAH-----GGGYGAGGAGGEGYGGGAGAGHGGGGGGGGGGGGGGGGGG 230
RESULT 12
S31415
glycine-rich protein GRP22 - rape
C:Species: Brassica napus (rape)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: S31415
R:Bergeron, D.; Boivin, R.; Baszczynski, C.L.; Bellemare, G.
submitted to the EMBL Data Library, August 1992
A:Description: Characterization and expression of a gene family encoding glycine-rich
A:Reference number: S31415
A:Accession: S31415
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-291 <BER>
A:Cross-references: EMBL:Z15045; NID:g17820; PIDN:CAA78762.1; PID:g17821
C:Superfamily: Phaseolus glycine-rich cell wall protein 1.8

Query Match 23.9% Score 130; DB 1; Length 291;
Best Local Similarity 50.0%; Pred. No. 4.3e-05;
Matches 31; Conservative 2; Mismatches 13; Indels 16; Gaps 3;

QY 18 GSGGKCAACALWLOCASGGTIGCGGAVACQNYRQFCRGGGGGGGGGGGGGG 73
Db 221 GSGAGAH-----GGYGAGG--AGEGFGGGGGGGGGGGGGGGGGGGGG 268
QY 74 GG 75
Db 269 GG 270

RESULT 13

F84797
hypothetical protein At2g37830 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: F84797
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487
A:Accession: F84797
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-106 <SPO>
A:Cross-references: GB:AE002093; NID:g4895201; PIDN:AA032788.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g37830
A:Map position: 2

Query Match 23.7% Score 129; DB 2; Length 106;
Best Local Similarity 44.4%; Pred. No. 2.3e-05;
Matches 28; Conservative 1; Mismatches 12; Indels 22; Gaps 2;

QY 14 ENKSGGLGKCAACALWLOCASGGTIGCGGAVACQNYRQFCRGGGGGGGGGGGGGG 73
Db 64 DGANGGGFG-----GDGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 101
QY 74 GGG 76
Db 102 GGG 104

RESULT 14

CIHUL
calpain (EC 3.4.22.17) small chain - human
N:Alternate names: calcium-activated neutral proteinase (CANP)
C:Species: Homo sapiens (man)
C:Date: 28-Dec-1987 #sequence_revision 28-Dec-1987 #text_change 16-Jul-1999
C:Accession: A26107; A23650
R:Miyake, S.; Emori, Y.; Suzuki, K.
Nucleic Acids Res. 14, 8805-8817, 1986
A:Title: Gene organization of the small subunit of human calcium-activated neutral prote
A:Reference number: A93648; MUID:87066759
A:Accession: A26107
A:Molecule type: DNA
A:Residues: 1-268 <MIY>
A:Cross-references: GB:M31502
R:Ohno, S.; Emori, Y.; Suzuki, K.
Nucleic Acids Res. 14, 5559, 1986
A:Title: Nucleotide sequence of a cDNA coding for the small subunit of human calcium-dep
A:Reference number: A93631; MUID:86286563
A:Accession: A23650
A:Molecule type: mRNA
A:Residues: 1-268 <ORN>
A:Cross-references: EMBL:X04106; NID:g35327; PIDN:CAA27726.1; PID:g35328
C:Genetics:
A:Gene: GDB:CAPN4

A:Cross-references: GDB:119752; OMIM:114170
A:Map position: 19pter-19qter
A:Introns: 70/2; 81/3; 111/3; 131/1; 152/3; 175/3; 202/1; 241/1; 260/3
C:Complex: heterodimer of L (large) and S (small) chains
C:Function:
A:Description: catalyzes the hydrolysis of peptides
A:Note: cleaves preferentially after tyrosine, methionine, or arginine residues and b
C:Superfamily: calpain small chain; calmodulin repeat homology
C:Keywords: calcium binding; cysteine proteinase; duplication; EF hand; heterodimer;
F:1-56/Domain: glycine-rich <GLY>
F:96-127/Domain: calmodulin repeat homology <EF1>
F:139-171/Domain: calmodulin repeat homology <EF2>
F:172-201/Domain: calmodulin repeat homology <EF3>
F:204-236/Domain: calmodulin repeat homology <EF4>
F:237-268/Domain: calmodulin repeat homology <EF5>

Query Match 23.7% Score 129; DB 1; Length 268;
Best Local Similarity 50.0%; Pred. No. 4.9e-05;
Matches 30; Conservative 0; Mismatches 18; Indels 12; Gaps 1;
QY 17 KSGSLGKCAACALWLOCASGGTIGCGGAVACQNYRQFCRGGGGGGGGGGGGGG 76
Db 9 KGGGG-----GGGGGGLGGLGNVLGSLISGAGGGGGGGGGGGGGGGGG 56

RESULT 15

WJHU2G
homeotic protein Hox B3 - human
N:Alternate names: homeotic protein Hox 2.7; homeotic protein Hox 2G
C:Species: Homo sapiens (man)
C:Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 22-Jun-1999
C:Accession: S07543; S15547; D37042
R:Acampora, D.; d'Esposito, M.; Faiella, A.; Pannese, M.; Migliaccio, E.; Morelli, F.
Nucleic Acids Res. 17, 10385-10402, 1989
A:Title: The human Hox gene family.
A:Reference number: S07541; MUID:90098876
A:Accession: S07543
A:Molecule type: mRNA
A:Residues: 1-431 <ACA>
A:Cross-references: EMBL:X16667; NID:g32379; PIDN:CAA34657.1; PID:g32380
R:Boncinelli, E.; Acampora, D.; Pannese, M.; d'Esposito, M.; Somma, R.; Gaudino, G.;
Genome 31, 745-756, 1989
A:Title: Differential expression of human Hox-2 genes along the anterior-posterior ax
A:Reference number: A37042; MUID:89378558
A:Accession: D37042
A:Molecule type: DNA
A:Residues: 188-253 <BON>
R:Giampaolo, A.; Acampora, D.; Zappavigna, V.; Pannese, M.; D'Esposito, M.; Care, A.;
Differentiation 40, 191-197, 1989
A:Title: Differential expression of human Hox-2 genes along the anterior-posterior ax
A:Reference number: A37042; MUID:89378558
A:Accession: D37042
A:Molecule type: DNA
A:Residues: 188-253 <GIA>
A:Cross-references: GB:X16175; NID:g32377; PIDN:CAA34297.1; PID:g930068
C:Genetics:
A:Gene: GDB:HOXB3
A:Map position: 17q21.3-17q21.3
A:Introns: 150/1
C:Superfamily: homeotic protein Hox B3; homeobox homology
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
F:152-178/Region: glycine-rich
F:189-245/Domain: homeobox homology <HOX>

Query Match 23.7% Score 129; DB 1; Length 431;
Best Local Similarity 58.7%; Pred. No. 7.3e-05;
Matches 27; Conservative 0; Mismatches 7; Indels 12; Gaps 2;
QY 35 SGGTI-GCGGAVACQNYRQFCRGGGGGGGGGGGGGGGGGGGGGGGGGGGG 79
Db 146 SPGTAEKGGGG-----GGGGGGGGGGGGGGGGGGGGGGGGGGGG 180

Thu Oct 24 14:40:01 2002

us-09-894-030-3.rpr

Page 6

Search completed: October 24, 2002, 09:29:44
Job time : 20.9781 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 09:28:26 ; Search time 56.2044 Seconds
(without alignments)
531.492 Million cell updates/sec

Title: US-09-894-030-3

Perfect score: 545

Sequence: 1 MEKLFKEVLEENQKSG.....DDFDLVVVKSKQDKITPQ 100

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1105779 seqs, 298721915 residues

Total number of hits satisfying chosen parameters: 1105779

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Pending Patents AA.New.*
1: /cgn2_6/ptodata/2/paa/PCT_NEW_COMB.pep.*
2: /cgn2_6/ptodata/2/paa/US05_NEW_COMB.pep.*
3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pep.*
4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pep.*
5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep.*
6: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep.*
7: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	165	30.3	303	5	US-09-791-537-31453
2	165	30.3	336	5	US-09-791-537-57429
3	147	27.0	333	5	US-09-791-537-96846
4	142.5	26.1	734	5	US-09-791-537-80458
5	142.5	26.1	918	5	US-09-791-537-74882
6	142.5	26.1	918	5	US-09-791-537-74905
7	142.5	26.1	919	5	PCT-US02-23913-36
8	142.5	26.1	919	5	US-09-791-537-85459
9	142.5	26.1	919	6	US-10-205-823-36
10	141.5	26.0	280	6	US-10-259-165-212
11	141	25.9	318	5	US-09-791-537-78475
12	139	25.5	335	5	US-09-791-537-22910
13	135.5	24.9	240	5	US-09-945-182-30
14	135	24.8	331	5	US-09-791-537-88976
15	134.5	24.7	101	6	US-10-182-995-20071
16	134.5	24.7	101	6	US-10-203-134-26311
17	134.5	24.7	101	6	US-10-203-136-26319
18	134.5	24.7	101	6	US-10-182-993-25731
19	134.5	24.7	101	6	US-10-182-997-18700
20	134.5	24.7	101	6	US-10-182-998-10357
21	134.5	24.7	101	6	US-10-203-135-25318
22	134.5	24.7	101	6	US-10-203-137-26550
23	134.5	24.7	101	6	US-10-203-138-10617
24	134.5	24.7	101	6	US-10-203-139-25454
25	134.5	24.7	447	5	US-09-791-537-64130
26	134.5	24.7	706	6	US-10-219-999-47643

27 134.5 24.7 1084 6 US-10-217-700-9 Sequence 9, Appli
28 134.5 24.7 1150 6 US-10-123-155-531 Sequence 531, App
29 134.5 24.7 1150 6 US-10-137-871-531 Sequence 531, App
30 134.5 24.7 1150 6 US-10-141-761-531 Sequence 531, App
31 134.5 24.7 1150 6 US-10-140-864-531 Sequence 531, App
32 134.5 24.7 1150 6 US-10-140-923-531 Sequence 531, App
33 134.5 24.7 1150 6 US-10-141-756-531 Sequence 531, App
34 134.5 24.7 1150 6 US-10-141-759-531 Sequence 531, App
35 134.5 24.7 1150 6 US-10-140-472-531 Sequence 531, App
36 134.5 24.7 1150 6 US-10-140-805-531 Sequence 531, App
37 134.5 24.7 1150 6 US-10-142-885-531 Sequence 531, App
38 134.5 24.7 1150 6 US-10-146-731-531 Sequence 531, App
39 134.5 24.7 1150 6 US-10-142-428-531 Sequence 531, App
40 134.5 24.7 1150 6 US-10-158-790-531 Sequence 531, App
41 134.5 24.7 1226 6 US-10-155-881-28229 Sequence 28229, A
42 133.5 24.5 241 6 US-10-155-881-27910 Sequence 27910, A
43 133.5 24.5 274 6 US-10-155-881-25833 Sequence 25833, A
44 133.5 24.5 1843 6 US-10-123-155-309 Sequence 309, App
45 133.5 24.5 1843 6 US-10-137-871-309 Sequence 309, App

ALIGNMENTS

RESULT 1

US-09-791-537-31453

; Sequence 31453, Application US/09791537

; GENERAL INFORMATION:

; APPLICANT: Bionomix, Inc.

; APPLICANT: Debe, Derek

; APPLICANT: Danzer, Joseph

; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME

; TITLE OF INVENTION: METHODS OF USE THEREOF

; FILE REFERENCE: 261/210

; CURRENT APPLICATION NUMBER: US/09/791,537

; CURRENT FILING DATE: 2001-02-22

; NUMBER OF SEQ ID NOS: 153055

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 31453

; LENGTH: 303

; TYPE: PRT

; ORGANISM: Populus balsamifera subsp

; US-09-791-537-31453

Query Match 30.3%; Score 165; DB 5; Length 303;

Best Local Similarity 42.4%; Pred. No. 2.8e-07;

Matches 42; Conservative 8; Mismatches 33; Indels 16; Gaps 5;

QY 4 LFKEVKLEENQKSGGLGKAQCAALWLCASGGTIG-----CGGAVACONYRQFCRG 58

DB 11 LFLSLVRGSAEQCGQAGDALCPG-GLCCSSYGCWCTTADYCGDG---CQSQ---CDGG 63

QY 59 GGGGGGGGGGGGGGGGMSKFDFFDLVVVKSKQDSKI 97

DB 64 GGGGGGGGGGGGGGGG---DGYLSDIIPESMFMDML 98

RESULT 2

US-09-791-537-57429

; Sequence 57429, Application US/09791537

; GENERAL INFORMATION:

; APPLICANT: Bionomix, Inc.

; APPLICANT: Debe, Derek

; APPLICANT: Danzer, Joseph

; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME

; TITLE OF INVENTION: METHODS OF USE THEREOF

; FILE REFERENCE: 261/210

; CURRENT APPLICATION NUMBER: US/09/791,537

; CURRENT FILING DATE: 2001-02-22

; NUMBER OF SEQ ID NOS: 153055

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 57429

; LENGTH: 336

```
; TYPE: PRT
; ORGANISM: Populus balsamifera subsp
US-09-791-537-57429

Query Match
Best Local Similarity 30.3%; Score 165; DB 5; Length 336;
Matches 42; Conservative 8; Mismatches 33; Indels 16; Gaps 5;

QY 4 LFEKVKLELENKQSGGLGKACAAALWLCQASGTTIG-----CGGAVACQNYRQFCRG 58
DB 11 LFLSLSVRGSAEQCGQAGDALCPG-GLCCSSYGCWGTADYCGDG---CQSQ---CDGG 63
QY 59 GGGGGGGGGGGGGGMSKFDFFDLVDVKVSKQDSKI 97
DB 64 GGGGGGGGGGGGGGGG-----DGLSDIIPESMFDDML 98

RESULT 3
US-09-791-537-96846
; Sequence 96846, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 96846
; LENGTH: 333
; TYPE: PRT
; ORGANISM: Oryza sativa
US-09-791-537-96846

Query Match
Best Local Similarity 27.0%; Score 147; DB 5; Length 333;
Matches 36; Conservative 2; Mismatches 16; Indels 12; Gaps 4;

QY 16 QKSGGLGKACAAALWLCQASGTTIG-----CGGAVACQNYRQFCRGCGGGGGGGGG 70
DB 25 QCGSQAGGALCPNC-LCCSSYGCWGSTDYCGDG---CQSQ---CDGCGGGGGGGGGGG 77
QY 71 GGGGGG 76
DB 78 GGGGGG 83

RESULT 4
US-09-791-537-80458
; Sequence 80458, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 80458
; LENGTH: 734
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-791-537-80458

Query Match
Best Local Similarity 26.1%; Score 142.5; DB 5; Length 734;
Matches 37; Conservative 3; Mismatches 22; Indels 25; Gaps 4;

; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-791-537-74882
; Sequence 74882, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 74882
; LENGTH: 917
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-791-537-74882

Query Match
Best Local Similarity 26.1%; Score 142.5; DB 5; Length 917;
Matches 37; Conservative 3; Mismatches 22; Indels 25; Gaps 4;

QY 12 ELENKQSGGLGKACAAALWLCQASG-----GTTCGGG---AVACQNYRQF----- 54
DB 198 KLENPLDYGSAAAAA---QCRYGDLASLHGAGAGPGSGSPSAASSSWHTLFTAERG 254
QY 55 -----CRGGGGGGGGGGGGGGGGGGGG 76
DB 255 QLYGPCGGGGGGGGGGGGGGGGGGGG 281

RESULT 5
US-09-791-537-74882
; Sequence 74882, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 74882
; LENGTH: 917
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-791-537-74882

Query Match
Best Local Similarity 26.1%; Score 142.5; DB 5; Length 917;
Matches 37; Conservative 3; Mismatches 22; Indels 25; Gaps 4;

QY 12 ELENKQSGGLGKACAAALWLCQASG-----GTTCGGG---AVACQNYRQF----- 54
DB 395 KLENPLDYGSAAAAA---QCRYGDLASLHGAGAGPGSGSPSAASSSWHTLFTAERG 441
QY 55 -----CRGGGGGGGGGGGGGGGGGGGG 76
DB 442 QLYGPCGGGGGGGGGGGGGGGGGGGG 468

RESULT 6
US-09-791-537-74905
; Sequence 74905, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 74905
; LENGTH: 918
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-791-537-74905

Query Match
Best Local Similarity 26.1%; Score 142.5; DB 5; Length 918;
Matches 37; Conservative 3; Mismatches 22; Indels 25; Gaps 4;

QY 12 ELENKQSGGLGKACAAALWLCQASG-----GTTCGGG---AVACQNYRQF----- 54
DB 382 KLENPLDYGSAAAAA---QCRYGDLASLHGAGAGPGSGSPSAASSSWHTLFTAERG 438
QY 55 -----CRGGGGGGGGGGGGGGGGGGGG 76
```



```
; TITLE OF INVENTION: GENES THAT ARE MODULATED BY POSTTRANSCRIPTIONAL GENE SILENCING
; FILE REFERENCE: 70030-NP
; CURRENT APPLICATION NUMBER: US/10/259,165
; CURRENT FILING DATE: 2002-09-26
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: US 60/369,327
; PRIOR FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; NUMBER OF SEQ ID NOS: 782
; SOFTWARE: PatentList.pl version 3.0.4 (C) 2001 Syngenta
; SEQ ID NO 212
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-259-165-212

Query Match      26.0%; Score 141.5; DB 6; Length 280;
Best Local Similarity 46.2%; Pred. No. 3.4e-05;
Matches 30; Conservative 3; Mismatches 15; Indels 17; Gaps 3;

QY 20 GLGKAQCAALWLQCAAGTIGCGGAVACQNYRQFCR-----GGGGGGGGGGGGGGG 73
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 25 GCOSNMCCSKWYCGTGDY-CGDG-----CRSGPCYGGGGGGGGGGGGGGG 73
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 74 GGGMS 78
   | | | |
Db 74 GSGVS 78

RESULT 11
US-09-791-537-78475
; Sequence 78475, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMB
; TITLE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 78475
; LENGTH: 318
; TYPE: PRT
; ORGANISM: Allium sativum
US-09-791-537-78475

Query Match      25.9%; Score 141; DB 5; Length 318;
Best Local Similarity 46.2%; Pred. No. 4.3e-05;
Matches 37; Conservative 5; Mismatches 22; Indels 16; Gaps 5;

QY 4 LPEYKLELENGKSGGLKQCAALWLQCAAGTIG-----CGGAVACQNYRQFCRG 58
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 11 LRF-----NSYAOQCGSQAGGALCSNR-LCCSKFYCGYSTDPYCGTG---CQSQ---CGG 59
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 59 GGGGGGGGGGGGGGGGGGMS 78
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 60 GGGGGGGGGGGGGGGGSGVA 79
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 12
US-09-791-537-22910
; Sequence 22910, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMB
; TITLE OF INVENTION: METHODS OF USE THEREOF
```

```
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22910
; LENGTH: 335
; TYPE: PRT
; ORGANISM: Elaeagnus umbellata
US-09-791-537-22910

Query Match      25.5%; Score 139; DB 5; Length 335;
Best Local Similarity 51.6%; Pred. No. 6.8e-05;
Matches 33; Conservative 3; Mismatches 16; Indels 12; Gaps 4;

QY 16 QKSGGLKQCAALWLQCAAGTIG-----CGGAVACQNYRQFCRGGGGGGGGGG 70
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 24 QCGQLGGALCSG-GLCCSQWGYCGNTDPYCGDG---CQSQ---CDGSGGGGGGGGGGG 76
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 71 GGGG 74
   | | | |
Db 77 GGGG 80

RESULT 13
US-09-945-182-30
; Sequence 30, Application US/09945182
; GENERAL INFORMATION:
; APPLICANT: Celeste, Anthony J.
; APPLICANT: Wozney, John
; APPLICANT: Rosen, Vicki A.
; APPLICANT: Wolfman, Neil
; APPLICANT: Thomsen, Gerald H.
; APPLICANT: Melton, Douglas A.
; TITLE OF INVENTION: TENDON-INDUCING COMPOSITIONS
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENETICS INSTITUTE, INC.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/945,182
; FILING DATE: 31-Aug-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/808,324
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Lazar, Steven R.
; REGISTRATION NUMBER: 32,618
; REFERENCE/DOCKET NUMBER: 5202-D
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617 498-8260
; TELEFAX: 617 876-5851
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 240 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 30:
US-09-945-182-30

Query Match      24.9%; Score 135.5; DB 5; Length 240;
Best Local Similarity 40.4%; Pred. No. 0.0001;
```

```

Matches 36: Conservative 5: Mismatches 25: Indels 23: Gaps 4:
QY 2 EKLFEVKLE-----ELENQKSGILG--KAQCAALWLQ-CASGGTTCGGGGAVACON 50
      |||::||:
Db 53 ESLFEIRAAQARAAAEPPDPGPGAGSKANLGGRRRQRTALAGTRGXGS----- 106
      |||::||:
QY 51 YRQFCRGGGGGGGGGGGGGGGGGGGGMSK 79
      |||::||:
Db 107 -----GGGGGGGGGGGGGGGGGGGGAGR 129
      |||::||:

```

RESULT 14
US-09-791-537-88976
; Sequence 88976, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS
; TITLE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791 537

```
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 88976
; LENGTH: 331
; TYPE: PRT
; ORGANISM: Mus musculus
; US-09-791-337-88976

Query Match      24.88;   Score 135;   DB 5;   Length 331;
Best Local Similarity 38.9%; Pred. No. 0.00015;
Matches 28; Conservative 11; Mismatches 19; Indels 14; Gaps 2;

Qy    34  ASGETIGCGGAVACQNYRQCRCGGGGGGGGGGGGGGGMSKFDFFDLDDVVVKVSKQ  93
       I::I I III |IIIII|IIII|IIII|IIII| :|: :|: :|:
Db     57  AAGGGGGAGGGGGG-----GGGGGGGAGGGAGGGGRSPVRELDMGAAERSRE  107

Qy    94  D-----SKITPQ 100
       :::::
Db     108  PGSPRLTEVSPSE 119
```

RESULT# 15
US-10-182-995-20071
; Sequence 20071, Application US/10182995
; GENERAL INFORMATION:
; APPLICANT: Molecular Dynamics, Inc.
; APPLICANT: Penn, Sharrion G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: ANALYSIS OF GENE EXPRESSION IN HUMAN HEART
; FILE REFERENCE: PB 0004 WO 1
; CURRENT APPLICATION NUMBER: US/10/182,995
; CURRENT FILING DATE: 2002-08-02
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 04 February 2000 (04.02.00)
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 26 May 2000 (26.05.00)
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 03 August 2000 (03.08.00)
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 03 October 2000 (03.10.00)
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 27 September 2000 (27.09.00)
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 21 September 2000 (21.09.00)
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 30 June 2000 (30.06.00)
; NUMBER OF SEQ ID NOS: 29119

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 09:24:01 : Search time 11.6788 Seconds
(without alignments)
331.536 Million cell updates/sec

Title: US-09-894-030-3
Perfect score: 545
Sequence: 1 MEKLFKEVKLEENKQSG.....DDFDLDVVKVKQDSKITPQ 100

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues
Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	165	30.3	303	1 CHIB_POPTR	P29031 populus tri
2	142.5	26.1	919	1 ANDR_HUMAN	P10275 homo sapien
3	135	24.8	331	1 SHX2_MOUSE	P70390 mus musculus
4	132.5	24.3	433	1 HXB3_MOUSE	P09026 mus musculus
5	132	24.2	331	1 SHX2_HUMAN	O60902 homo sapien
6	130	23.9	165	1 GRP1_ORYSA	P25074 oryza sativ
7	129	23.7	268	1 CANS_BOVIN	P04632 homo sapien
8	129	23.7	431	1 HXB3_HUMAN	P14651 homo sapien
9	128	23.5	214	1 GRP2_NICSY	P27484 nicotiana s
10	127	23.3	384	1 GRP1_PETHY	P09789 petunia hyb
11	126.5	23.2	911	1 ANDR_PANTR	O97775 pan troglod
12	126	23.1	416	1 R23B_MOUSE	P54728 mus musculus
13	126	23.1	476	1 EVX2_HUMAN	Q03828 homo sapien
14	125.5	23.0	404	1 CAZ_DROME	Q27294 drosophila
15	125	22.9	151	1 GOF7_MOUSE	P43029 mus musculus
16	125	22.9	266	1 CANS_PIG	P04574 sus scrofa
17	124.5	22.8	321	1 FBRL_HUMAN	P22087 homo sapien
18	124	22.8	201	1 GR2B_ARATH	Q38896 arabidopsis
19	124	22.8	322	1 PUR_HUMAN	Q00577 homo sapien
20	124	22.8	1627	1 TF2B_CHICK	O42131 gallus gall
21	123.5	22.7	321	1 PUR_MOUSE	P42669 mus musculus
22	123	22.6	280	1 CHIA_MAIZE	P29022 zea mays (m
23	122	22.4	520	1 WASP_MOUSE	P70315 mus musculus
24	121.5	22.3	268	1 EP34_HCMVA	P16768 human cytom
25	121.5	22.3	316	1 LORI_HUMAN	P23490 homo sapien
26	121.5	22.3	481	1 LORI_MOUSE	P18165 mus musculus
27	121.5	22.3	684	1 EP84_HCMVA	P17151 human cytom
28	121	22.2	56	1 LANS_BACSU	P10946 bacillus su
29	120.5	22.1	199	1 AC22_TENMO	P28968 tenebrio mo
30	120.5	22.1	266	1 CANS_RABIT	P06813 oryctolagus
31	120.5	22.1	332	1 SIX3_HUMAN	O95343 homo sapien
32	120.5	22.1	440	1 FXGA_CHICK	Q98937 gallus gall
33	120.5	22.1	458	1 K1CM_HUMAN	P13646 homo sapien

34	120	22.0	443	1 OC3N_HUMAN	P20265 homo sapien
35	120	22.0	445	1 OC3N_MOUSE	P31360 mus musculus
36	120	22.0	533	1 ZIC2_HUMAN	O95409 homo sapien
37	119.5	21.9	252	1 GRP1_PHAVU	P10495 phaseolus v
38	119.5	21.9	694	1 FZD8_HUMAN	O9H461 homo sapien
39	119.5	21.9	1380	1 DD9_MOUSE	O70133 mus musculus
40	119	21.8	206	1 TWST_MOUSE	P26687 mus musculus
41	119	21.8	495	1 BRN1_MOUSE	P31361 mus musculus
42	119	21.8	497	1 BRN1_RAT	O63262 rattus norv
43	119	21.8	500	1 BRN1_HUMAN	P20264 homo sapien
44	118	21.7	1093	1 PER_DROWI	Q03297 drosophila
45	117.5	21.6	263	1 CANS_BOVIN	P13135 bos taurus

ALIGNMENTS

RESULT 1
CHIB_POPTR
ID CHIB_POPTR STANDARD; PRT; 303 AA.
AC P29031;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Acidic endochitinase WIN6.2B precursor (EC 3.2.1.14).
OS Populus trichocarpa (Western balsam poplar).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Malpighiales; Salicaceae; Populus.
OX NCBI_TaxID=3694;
RN [1]
RP SEQUENCE FROM N. A.
RX MEDLINE=92003678; Pubmed=1912489;
RA Davis J.M., Clarke H.R.G., Bradshaw H.D. Jr., Gordon M.P.;
RT "Populus chitinase genes: structure, organization, and similarity of
RT translated sequences to herbaceous plant chitinases.";
RL Plant Mol. Biol. 17:631-639(1991).
CC -!- FUNCTION: THIS PROTEIN FUNCTIONS AS A DEFENSE AGAINST CHITIN
CC CONTAINING FUNGAL PATHOGENS.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of the 1,4-beta-linkages of N-
CC acetyl-D-glucosamine polymers of chitin.
CC -!- INDUCTION: By wounding.
CC -!- SIMILARITY: BELONGS TO CHITINASE CLASS IA WHICH IS SIMILAR TO
CC CLASS IB IN THE CATALYTIC PART BUT WHICH INCLUDE A N-TERMINAL
CC CHITIN-BINDING LECTIN DOMAIN (BELONGS TO FAMILY 19 OF GLYCOSYL
CC HYDROLASES).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; X59995; CAA42612.1; --
CC HSP; P02877; 1HEV
DR InterPro; IPR001002; Chitin_bind.
DR InterPro; IPR000726; Glyco_hydro_19.
DR Pfam; PF00187; chitin_binding; 1.
DR Pfam; PF00182; Glyco_hydro_19; 2.
DR PRINTS; PR00451; CHITINBINDG.
DR ProDom; PD000574; Glyco_hydro_19; 1.
DR ProDom; PD000609; Chitin_bind; 1.
DR SMART; SM00270; ChEBDI; 1.
DR PROSITE; PS00026; CHITIN_BINDING; 1.
DR PROSITE; PS00773; CHITINASE_19; 1.
DR PROSITE; PS00774; CHITINASE_19; 2; 1.
KW Hydrolase; Glycosidase; Chitin degradation; Chitin-binding; Signal;
FT SIGNAL 1 21 BY SIMILARITY.
FT CHAIN 22 303 ACIDIC ENDOCHITINASE WIN6.2B.
FT DOMAIN 22 61 CHITIN-BINDING (BY SIMILARITY).

```
FT DOMAIN 62 81 POLY-GLY (SPACER).
FT DOMAIN 82 303 CHITINASE.
FT DISULFID 24 39 BY SIMILARITY.
FT DISULFID 33 45 BY SIMILARITY.
FT DISULFID 38 52 BY SIMILARITY.
FT DISULFID 56 60 BY SIMILARITY.
SQ SEQUENCE 303 AA; 32208 MW; 0E0D011DFA5CCBB8 CRC64;

Query Match 30.3%; Score 165; DB 1; Length 303;
Best Local Similarity 42.4%; Pred. No. 3e-08;
Matches 42; Conservative 8; Mismatches 33; Indels 16; Gaps 5;

QY 4 LFEVKLEELLENQKSGLGKACQAAALWLCASGGTIG-----CGGGAACQNYQFQRRGG 58
Db 11 LFLSLVRGSAEOGQGGAGDALCPG-GLCCSSYGCWCTTADYCGDG---CQSQ---CDGG 63
QY 59 GGGGGGGGGGGGGGGGMSKFDFFDLDVVKVSKQDSKI 97
Db 64 GGGGGGGGGGGGGGGG-----DGLSLDIIPESFDDML 98

RESULT 2
ANDR_HUMAN
ID ANDR_HUMAN STANDARD; PRT; 919 AA.
AC P10275;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DE 01-MAR-2002 (Rel. 41, Last annotation update)
DE Androgen receptor (Dihydrotestosterone receptor).
GN AR OR NR3C4 OR DHTR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89112208; PubMed=3216866;
RA Lubahn D.B., Joseph D.R., Sar M., Tan J., Higgs H.N., Larson R.E.,
RA French F.S., Wilson E.M.;
RT "The human androgen receptor: complementary deoxyribonucleic acid
RT cloning, sequence analysis and gene expression in prostate.";
RL Mol. Endocrinol. 2:1265-1275(1988).
RN [2]
RP SEQUENCE FROM N.A., AND VARIANT CAIS MET-866.
RX MEDLINE=90083302; PubMed=2594783;
RA Lubahn D.B., Brown T.R., Simental J.A., Higgs H.N., Migeon C.J.,
RA Wilson E.M., French F.S.;
RT "Sequence of the intron/exon junctions of the coding region of the
RT human androgen receptor gene and identification of a point mutation
RT in a family with complete androgen insensitivity.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:9534-9538(1989).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=90258935; PubMed=2342476;
RA Govindan M.V.;
RT "Specific region in hormone binding domain is essential for hormone
RT binding and trans-activation by human androgen receptor.";
RL Mol. Endocrinol. 4:417-427(1990).
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=Prostate;
RX MEDLINE=89017168; PubMed=3174628;
RA Chang C., Kokontis J., Liao S.;
RT "Structural analysis of complementary DNA and amino acid sequences of
RT human and rat androgen receptors.";
RL Proc. Natl. Acad. Sci. U.S.A. 85:7211-7215(1988).
RN [5]
RP SEQUENCE FROM N.A.
RC TISSUE=Prostate;
RX MEDLINE=89098909; PubMed=2911578;
RA Tilley W.D., Marcelli M., Wilson J.D., McPhaul M.J.;
RT "Characterization and expression of a cDNA encoding the human androgen
RT receptor.";
```

```
RL Proc. Natl. Acad. Sci. U.S.A. 86:327-331(1989).
RN [6]
RP SEQUENCE FROM N.A.
RC TISSUE=Prostate;
RX MEDLINE=91155943; PubMed=2293020;
RA Marcelli M., Tilley W.D., Wilson C.M., Griffin J.E., Wilson J.D.,
RA McPhaul M.J.;
RT "Definition of the human androgen receptor gene structure permits the
RT identification of mutations that cause androgen resistance: premature
RT termination of the receptor protein at amino acid residue 588 causes
RT complete androgen resistance.";
RL Mol. Endocrinol. 4:1105-1116(1990).
RN [7]
RP SEQUENCE OF 189-919 FROM N.A.
RX MEDLINE=88178111; PubMed=3353726;
RA Chang C., Kokontis J., Liao S.;
RT "Molecular cloning of human and rat complementary DNA encoding
RT androgen receptors.";
RL Science 240:324-326(1988).
RN [8]
RP SEQUENCE OF 468-919 FROM N.A.
RX MEDLINE=88240407; PubMed=3377788;
RA Trapman J., Klaassen P., Kuiper G.G.J.M., van der Korput J.A.G.M.,
RA Faber P.W., van Rooij H.C.J., Geurts van Kessel A., Voorhorst M.M.,
RA Mulder E., Brinkmann A.O.;
RT "Cloning, structure and expression of a cDNA encoding the human
RT androgen receptor.";
RL Biochem. Biophys. Res. Commun. 153:241-248(1988).
RN [9]
RP INTERACTION WITH RAN.
RX MEDLINE=99329028; PubMed=10400640;
RA Hsiao P.-W., Lin D.-L., Nakao R., Chang C.;
RT "The linkage of Kennedy's neuron disease to ARA24, the first
RT identified androgen receptor polyglutamine region-associated
RT coactivator.";
RL J. Biol. Chem. 274:20229-20234(1999).
RN [10]
RP POLYMORPHISM OF POLY-GLN REGION.
RX MEDLINE=92220629; PubMed=1561105;
RA Sledzews H.F.B.M., Oostra B.A., Brinkmann A.O., Trapman J.;
RT "Trinucleotide repeat polymorphism in the androgen receptor gene
RT (AR).";
RL Nucleic Acids Res. 20:1427-1427(1992).
RN [11]
RP POLYMORPHISM OF POLY-GLY REGION.
RC TISSUE=Blood;
RA Lu J., Danielsen M.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
RN [12]
RP POLYMORPHISM OF POLY-GLN REGION.
RX MEDLINE=97250535; PubMed=9096391;
RA Giovannucci E., Stampfer M.J., Krithivas K., Brown M., Brufsky A.,
RA Talcott J., Hennekens C.H., Kantoff P.W.;
RT "The CAG repeat within the androgen receptor gene and its
RT relationship to prostate cancer.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:3320-3323(1997).
RN [13]
RP ERRATUM.
RA Giovannucci E., Stampfer M.J., Krithivas K., Brown M., Dahl D.,
RA Brufsky A., Talcott J., Hennekens C.H., Kantoff P.W.;
RL Proc. Natl. Acad. Sci. U.S.A. 94:8272-8272(1997).
RN [14]
RP REVIEW ON VARIANTS.
RX MEDLINE=93092459; PubMed=1458719;
RA Pinsky L., Trifiro M.A., Kaufman M., Beitel L.K., Mhatre A.,
RA Kazemi-Esfarjani P., Sabbaghian N., Lumbruso R., Alvarado C.,
RA Vasilion M., Gottlieb B.;
RT "Androgen resistance due to mutation of the androgen receptor.";
RL Clin. Invest. Med. 15:456-472(1992).
RN [15]
RP REVIEW ON VARIANTS AIS.
RX MEDLINE=93339360; PubMed=8339746;
RA Brown T.R., Scherer P.A., Chang Y.-T., Migeon C.J., Ghirri P.,
```


Db 142 LKNSFGTAEGCGGG-----GGGCGGGCGGGSSGGGGGGGGGDK 183

RESULT 5
SHX2_HUMAN PRT; 331 AA.
ID SHX2_HUMAN STANDARD; O60465; O60467;
AC O60502; O60903; O60465; O60467;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE DE SHOT status homeobox protein 2 (Paired-related homeobox protein
SHORT) (Homeobox protein Ogl12X).
GN SHOX2 OR SHOT OR OG12X.
OS Homo sapiens (Human),
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxId=9606;
[1]
SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RP TISSUE=Fibroblast;
RC MEDLINE=98151525; PubMed=9482898;
RX Blaschke R.J., Monaghan A.P., Schiller S., Schechinger B., Rao E.,
RA Padilla-Nash H., Ried T., Rappold G.A.;
RT "SHOT, a SHOX-related homeobox gene, is implicated in craniofacial,
brain, heart, and limb development.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:2406-2411(1998).
RN [2]
SEQUENCE FROM N.A.
RP TISSUE=Muscle;
RC Strausberg R.;
RA Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
SEQUENCE OF I16-331 FROM N.A. (ISOFORM 2).
RP TISSUE=Craniofacial;
RC MEDLINE=98133920; PubMed=9466998;
RX Semina E.V., Reiter R.S., Murray J.C.;
RA "A new human homeobox gene Og12x is a member of the most conserved
homeobox gene family and is expressed during heart development in
mouse.";
RL Hum. Mol. Genet. 7:415-422(1998).
RN [-] FUNCTION: May be a growth regulator and have a role in specifying
neural systems involved in processing somatosensory information,
as well as in face and body structure formation.
CC [-] SUBCELLULAR LOCATION: Nuclear.
CC [-] ALTERNATIVE PRODUCTS: 2 isoforms; 1/SHOX2A/SHOTa (shown here) and
2/SHOX2B/SHOTb/OG12XB; are produced by alternative splicing.
CC [-] TISSUE SPECIFICITY: Expressed in heart, skeletal muscle, liver,
lung, bone marrow fibroblast, pancreas and placenta.
CC [-] DEVELOPMENTAL STAGE: Expressed during craniofacial development as
well as in heart.
CC [-] SIMILARITY: BELONGS TO THE PAIRED FAMILY OF HOMEBOX PROTEINS.
CC ["BCOID" SUBFAMILY.
CC [-] SIMILARITY: CONTAINS 1 OAR DOMAIN.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announcement/
or send an email to license@isb-sib.ch).

EMBL; AJ002367; CAA05341.1; ALT_INIT.
DR EMBL; AJ002368; CAA05342.1; ALT_INIT.
DR EMBL; BC008829; AAAH08829.1; .
DR EMBL; AF022654; AAC39662.1; ALT_INIT.
DR EMBL; AF023203; AAC39663.1; .
DR HSSP; P06601; IFUL.
DR MIM; 602504; .
DR InterPro; IPR000047; HTH_repressor.
DR InterPro; IPR001356; Homeobox.
DR InterPro; IPR003654; OAR_domain.
DR Pfam; PF000046; homeobox; 1.


```

Db 145 GGGYG-----GGGGGGGCKGCGESGHFARDCSQSGGGGGGRFGGGGGG 191
QY 73 GGGGMSK 79
    |||||
Db 192 GGGGCYK 198

RESULT 10
GRPI_PETHY STANDARD; PRT; 384 AA.
AC P09789;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-AUG-1992 (Rel. 23, Last annotation update)
DE Glycine-rich cell wall structural protein 1 precursor.
GN GRP-1.
OS Petunia hybrida (Petunia).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Petunia.
OX NCBI_TaxID=4102;
RN [1]
RN SEQUENCE FROM N.A.
RA Condit C.M., Meagher R.B.;
RT "A gene encoding a novel glycine-rich structural protein of petunia.";
RL Nature 373:178-181(1986).
CC -1- FUNCTION: RESPONSIBLE FOR PLASTICITY OF THE CELL WALL (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: Cell wall (Potential).
CC -1- MISCELLANEOUS: THIS PROTEIN CONTAINS 67% GLYCINE RESIDUES.
CC FORMING A BETA-PLATED SHEET COMPOSED OF 8 ANTI-PARALLEL STRANDS.
CC -1- MISCELLANEOUS: THE GLYCINE-RICH REGION IS COMPRISED OF TWO RELATED
CC FAMILIES OF REPEATS, F1 AND F2, EACH REPEAT CONTAINING ABOUT 40
CC AA.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; X04335; CAA27866.1; -.
DR EMBL; A26099; A26099.
DR HSSP; P30129; 4DPV.
KW Cell wall; Structural protein; Repeat; Signal.
FT SIGNAL 1 27
FT CHAIN 28 384 GLYCINE-RICH CELL WALL STRUCTURAL
FT PROTEIN 1.
FT DOMAIN 41 384 GLY-RICH.
FT SEQUENCE 384 AA; 28777 MW; C8541C549417D18C CRC64;

Query Match 23.3%; Score 127; DB 1; Length 384;
Best Local Similarity 47.5%; Pred. No. 9.4e-05;
Matches 29; Conservative 2; Mismatches 6; Indels 24; Gaps 3;

QY 18 GSGLGKQAACALWLQCSGTTI--GGGGGAVACNQYRQFCRGGGGGGGGGGGGGGG 75
Db 232 GGGFG-----AGGGVGGVGGGAA-----GGGGGGGGGGGGGLGG 269
QY 76 G 76
Db 270 G 270

RESULT 11
ANDR_PANTR STANDARD; PRT; 911 AA.
AC O97775;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)

```

```

DE Androgen receptor (Dihydrotestosterone receptor).
GN AR OR NR3C4.
OS Pan troglodytes (Chimpanzee).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
OX NCBI_TaxID=9598;
RN [1]
RN SEQUENCE FROM N.A.
RA MEDLINE-98404153; PubMed-9732460;
RA Choong C.S., Kempainen J.A., Wilson E.M.;
RT "Evolution of the primate androgen receptor: a structural basis for
RT disease.";
RL J. Mol. Evol. 47:334-342(1998).
CC -1- FUNCTION: THE STEROID HORMONES AND THEIR RECEPTORS ARE INVOLVED IN
CC THE REGULATION OF EUKARYOTIC GENE EXPRESSION AND AFFECT CELLULAR
CC PROLIFERATION AND DIFFERENTIATION IN TARGET TISSUES.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
CC NR3 SUBFAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; U94177; AAC73048.1; -.
DR HSSP; P06536; 1GDC.
DR InterPro; IPR001103; Androgen_recep.
DR InterPro; IPR000536; Hormone_rec_lig.
DR InterPro; IPR001628; zf-C4.
DR Pfam; PF02166; Androgen_recep; 1.
DR Pfam; PF00104; hormone_rec; 1.
DR Pfam; PF00105; zf-C4; 1.
DR PRINTS; PR00047; STROIDFINGER.
DR SMART; SM00430; HOL1; 1.
DR SMART; SM00399; Znf_C4; 1.
DR PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW Zinc-finger; Steroid-binding.
FT DOMAIN 1 549 MODULATING (BY SIMILARITY).
FT DNA_BIND 551 616 NUCLEAR RECEPTOR-TYPE.
FT ZN_FING 551 571 C4-TYPE.
FT ZN_FING 587 611 C4-TYPE.
FT DOMAIN 682 911 LIGAND-BINDING.
FT DOMAIN 57 78 POLY-GLN.
FT DOMAIN 84 88 POLY-GLN.
FT DOMAIN 192 196 POLY-GLN.
FT DOMAIN 371 380 POLY-PRO.
FT DOMAIN 395 401 POLY-ALA.
FT DOMAIN 448 464 POLY-GLY.
FT SEQUENCE 911 AA; 98402 MW; 601B9BD4E697DAA4 CRC64;

Query Match 23.2%; Score 126.5; DB 1; Length 911;
Best Local Similarity 37.4%; Pred. No. 0.00021;
Matches 34; Conservative 6; Mismatches 26; Indels 25; Gaps 4;

QY 12 ELENQKSGGLGKQAACALWLQCSG-----GTTCGGG-----AVACNQYRQF----- 54
Db 385 KLENPLDYGSAAAAA---QCRYGLASLHGAGAGPGSGPSAAASSWHTLFTAE 441
QY 55 -----CRGGGGGGGGGGGGGGGMSKF 80
Db 442 QLYGPGCGGGGGGGGGGGGGGGAAGAVAPY 472

RESULT 12
R23B_MOUSE STANDARD; PRT; 416 AA.
ID R23B_MOUSE

```



```

RX MEDLINE=87175568; PubMed=3031652;
RT Haynes S.R., Rebert M.L., Mozer B.A., Forquignon F., Dawid I.B.;
RA 'pen repeat sequences are GGN clusters and encode a glycine-rich
RT domain in a Drosophila cDNA homologous to the rat helix destabilizing
RT protein.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:1819-1823(1987).
CC -!- FUNCTION: MAY PARTICIPATE IN A FUNCTION COMMON TO THE EXPRESSION
CC OF MOST GENES TRANSCRIBED BY RNA POLYMERASE II.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; 1 AND 2 (SHOWN HERE);
CC -!- MAY BE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: UBICUITOUS. ENRICHED IN THE BRAIN AND CENTRAL
CC NERVOUS SYSTEM DURING EMBRYOGENESIS. ENRICHED IN THE ADULT HEAD.
CC EMBRYOS CONTAIN BOTH TYPE 1 AND TYPE 2 ISOFORMS, WHEREAS LATER IN
CC DEVELOPMENT (HEADS AND TORSOS) ONLY THE TYPE 2 ISOFORM IS
CC DETECTED.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED IN THE DEVELOPING EMBRYO FROM THE
CC EARLIEST STAGES OF CELLULARIZATION AND IS SUBSEQUENTLY FOUND IN
CC MANY CELL TYPES.
CC -!- MISCELLANEOUS: 'CABEZA' MEANS 'HEAD' IN SPANISH.
CC -!- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
CC -!- SIMILARITY: CONTAINS 1 RANBP2-TYPE ZINC FINGER.
CC -!- SIMILARITY: BELONGS TO THE TET FAMILY OF RNP PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U13178; AA86955.1; -.
DR EMBL; L37083; AAC41563.1; -.
DR EMBL; AE003501; AAF48578.1; -.
DR EMBL; M15765; AAA70425.1; -.
DR FlyBase; FBgn0011571; caz.
DR InterPro; IPR000504; Rnf.
DR InterPro; IPR001876; Znf-RanBP.
DR Pfam; PF00076; rrm; 1.
DR Pfam; PF00641; zf-RanBP; 1.
DR SMART; SM00360; RRM; 1.
DR SMART; SM00547; Znf_RB2; 1.
DR PROSITE; PS50102; RRM; 1.
DR PROSITE; PS00030; RRM_RNP_1; FALSE_NEG.
DR PROSITE; PS01358; ZF_RANBP2_1; 1.
DR PROSITE; PS01099; ZF_RANBP2_2; 1.
DR Nuclear protein; Zinc-finger; Metal-binding; RNA-binding;
DR Alternative splicing.
DR DOMAIN 42 111 GLY-RICH.
DR DOMAIN 119 205 RNA-BINDING (RRM).
DR DOMAIN 212 275 GLY-RICH.
DR ZN_FING 280 309 RANBP2-TYPE.
DR DOMAIN 312 391 GLY-RICH.
DR VARSPIC 4 47 MISSING (IN ISOFORM 1).
DR CONFLICT 39 41 PNY -> LFI (IN REF. 4).
DR CONFLICT 92 92 P -> H (IN REF. 3).
DR CONFLICT 108 108 G -> GG (IN REF. 3).
DR CONFLICT 253 258 MISSING (IN REF. 3).
DR CONFLICT 283 283 D -> E (IN REF. 4 AND 5).
DR CONFLICT 389 398 DGGPMRDGG -> MVDQKRWIS (IN REF. 4).
DR SEQUENCE 404 AA; 39141 MW; 7062A0446BEA5984 CRC64;
Query Match 23.0%; Score 125.5; DB 1; Length 404;
Best Local Similarity 43.4%; Pred. No. 0.00013;
Matches 33; Conservative 3; Mismatches 13; Indels 27; Gaps
QY 7 EVKLELEN--QKSGSLGKAOCALWLCQASGGTICGGGGAVACQNYRQFCRGGGGGG 64
: : : : : | | | | | : : : : : | | | | |
Db 199 KVSIAQRQNNWNKNGGGGGG-----GGGRGGGFGGR-----RGGGGGGG 237
: : : : : | | | | |
QY 65 GGGGGG---GGGGG 76
|||||

```

'Thu Oct 24 14:40:01 2002

us-09-894-030-3.rsp

Db 238 GGGGGGRDRGGGGG 253

```
RESULT 15
GDF7_MOUSE
ID GDF7_MOUSE STANDARD; PRT; 151 AA.
AC P43029;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Growth/differentiation factor 7 precursor (GDF-7) (Fragment).
GN GDF7 OR GDF-7.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C; TISSUE=Liver;
RX MEDLINE=94195427; PubMed=8145850;
RA Storm E.E., Huynh T.V., Copeland N.G., Jenkins N.A., Kingsley D.M.,
RA Lee S.-J.;
RT "Limb alterations in brachypodism mice due to mutations in a new
member of the TGF-beta-superfamily.";
RL Nature 368:639-643(1994).
CC -1- SUBUNIT: HOMODIMER; DISULFIDE-LINKED (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE TGF-BETA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U08339; AAA18780.1; -.
DR HSP; P12643; 3BMP.
DR MGD; MGI:95690; Gdf7.
DR InterPro; IPR001839; TGF-beta.
DR Pfam; PF00019; TGF-beta; 1.
DR ProDom; PD000357; TGF-beta; 1.
DR SMART; SM00204; TGF-beta; 1.
DR PROSITE; PS00250; TGF-BETA_1; 1.
KW Growth factor; Cytokine; Glycoprotein.
FT NON_TER 1 1
FT PROPEP <1 5 POTENTIAL.
FT CHAIN 6 151 GROWTH/DIFFERENTIATION FACTOR 7.
FT DISULFID 50 116 BY SIMILARITY.
FT DISULFID 79 148 BY SIMILARITY.
FT DISULFID 83 150 BY SIMILARITY.
FT DISULFID 115 115 INTERCHAIN (BY SIMILARITY).
FT DOMAIN 1 5 POLY-ARG.
FT DOMAIN 16 41 POLY-GLY.
FT SEQUENCE 151 AA; 15697 MW; 0E496AACB5827759 CRC64;
```

```
Query Match 22.9%; Score 125; DB 1; Length 151;
Best Local Similarity 54.3%; Pred. No. 6.7e-05;
Matches 25; Conservative 1; Mismatches 8; Indels 12; Gaps 1;

QY 34 ASGGTTCGGGAVACQNYRQFCRGGGGGGGGGGGGGGGGGGGGGMSK 79
Db 7 ALAGTRGQGS-----GGGGGGGGGGGGGGGGGGGGGAGR 40
```

Search completed: October 24, 2002, 09:28:20
Job time : 14.6788 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 09:24:21 : Search time 32.1168 Seconds
(without alignments)
538.643 Million cell updates/sec

Title: US-09-894-030-3
Perfect score: 545
Sequence: 1 MEKLFKEVKLEELNKGSG.....DDFDLDVVKVSKDQSKITPO 100

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	304	55.8	56	9 O64033	O64033 bacterioph
2	304	55.8	56	16 O34781	O34781 bacillus su
3	151	27.7	385	5 Q93424	Q93424 caenorhabdi
4	147	27.0	333	10 Q42992	Q42992 oryza sativ
5	142.5	26.1	531	4 Q9BZG5	Q9bzg5 homo sapien
6	142.5	26.1	539	4 Q9NUA2	Q9nu2 homo sapien
7	142.5	26.1	542	4 Q9BZG6	Q9bzg6 homo sapien
8	142.5	26.1	544	4 Q9BZG7	Q9bzg7 homo sapien
9	141	25.9	221	10 O65514	O65514 arabidopsis
10	141	25.9	318	10 Q38777	Q38777 allium sati
11	139.5	25.6	233	16 Q92NU7	Q92nu7 rhizobium m
12	139.5	25.6	393	5 Q18880	Q18880 caenorhabdi
13	139	25.5	335	10 O65330	O65330 elaeagnus u
14	138	25.3	207	10 Q43522	Q43522 lycopersico
15	137	25.1	697	5 Q9GRW7	Q9grw7 drosophila
16	136.5	25.0	718	12 Q91TR1	Q91tr1 tupala herp

17	135	24.8	163	5	Q95UX4	Q95ux4 drosophila
18	135	24.8	165	5	Q9GP44	Q9gp44 drosophila
19	135	24.8	165	5	Q95NR6	Q95nr6 drosophila
20	134.5	24.7	165	5	Q95UX2	Q95ux2 drosophila
21	134.5	24.7	192	16	Q92P87	Q92p87 rhizobium m
22	134.5	24.7	447	13	O73628	O73628 anolis caro
23	134.5	24.7	452	10	Q9LW52	Q9lw52 arabidopsis
24	134.5	24.7	1084	10	Q43010	Q43010 oryza sativ
25	133.5	24.5	441	11	Q99MY1	Q99my1 mus musculu
26	133	24.4	106	5	Q9GP80	Q9gp80 drosophila
27	132	24.2	132	10	Q943G4	Q943g4 oryza sativ
28	131.5	24.1	163	5	Q95UW9	Q95uw9 drosophila
29	130.5	23.9	162	5	Q95UX5	Q95ux5 drosophila
30	130.5	23.9	164	5	Q95NP2	Q95np2 drosophila
31	130.5	23.9	175	10	Q9LSN6	Q9lsn6 arabidopsis
32	130	23.9	159	5	Q95UW1	Q95uw1 drosophila
33	130	23.9	161	5	Q95UX3	Q95ux3 drosophila
34	130	23.9	163	5	Q95NU6	Q95nu6 drosophila
35	130	23.9	168	5	Q95UW8	Q95uw8 drosophila
36	130	23.9	255	10	Q9SIH2	Q9sih2 arabidopsis
37	130	23.9	291	10	Q93337	Q93337 brassica na
38	129.5	23.8	738	5	O02402	O02402 pinctada fu
39	129	23.7	106	10	Q9SHA4	Q9sha4 arabidopsis
40	129	23.7	155	5	Q9GP74	Q9gp74 drosophila
41	129	23.7	155	5	Q9GND8	Q9gnd8 drosophila
42	129	23.7	321	6	Q9MYX6	Q9myx6 ovio aries
43	129	23.7	322	4	Q96EW0	Q96ew0 homo sapien
44	129	23.7	337	11	Q9ET75	Q9et75 rattus norv
45	129	23.7	698	5	Q9GRX4	Q9grx4 drosophila

ALIGNMENTS

RESULT 1

O64033 ID O64033 PRELIMINARY; PRT; 56 AA.
AC O64033;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PUTATIVE LIPOPROTEIN.
GN YOLG.
OS Bacteriophage SPBc2.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.
OX NCBI_TaxID=66797;
RN [1]
RP SEQUENCE FROM N.A.
RA Lazarevic V., Duesterhoeft A., Soldo B., Hilbert H., Mael C.,
RA Karamata D.;
RT "The complete nucleotide sequence of the Bacillus subtilis SPbetac2
RT prophage.";
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF020713; AAC12992.1; -
KW Lipoprotein.
SQ SEQUENCE 56 AA; 5982 MW; 79EC0BF822F9F4C0 CRC64;

Query Match 55.8%; Score 304; DB 9; Length 56;
Best Local Similarity 100.0%; Pred. No. 5.8e-26;
Matches 56; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MEKLFKEVKLEELNKGSGGLGKAQCAALWLQCAASGGTICGGGAVACQNYRPCR 56
|||||
DB 1 MEKLFKEVKLEELNKGSGGLGKAQCAALWLQCAASGGTICGGGAVACQNYRPCR 56
|||||

RESULT 2

O34781 ID O34781 PRELIMINARY; PRT; 56 AA.
AC O34781;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)

Query Match 27.08; Score 147; DB 10; Length 333;
 Best Local Similarity 54.58; Pred. No. 4.3e-08;
 Matches 36; Conservative 2; Mismatches 16; Indels 12; Gaps 4;
 QY 16 QKSGSLGKAQCAALW-LQCAAGTTIG-----CGGGAVACQNYRQFCRGGGGGGGGGGGGG 70
 DB 25 QCGSQAGGALCPNC-LCCSSYWGCGSTSDYCGDG---CQSQ---CDGCGGGGGGGGGGGG 77
 QY 71 GGGGGG 76
 DB 78 GGGGGG 83

RESULT 5

Q9BZG5 PRELIMINARY; PRT; 531 AA.
 AC Q9BZG5;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE ANDROGEN RECEPTOR (FRAGMENT).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Magklara A., Diamandis E.P.;
 RT "Poly-Q and poly-G repeats in the androgen receptor of the breast
 cancer cell line MCF-7";
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF321917; AAK09426.1; -;
 DR InterPro: IPR001103; Androgen_recep.
 DR Pfam: PF02166; Androgen_recep; 1.
 KW Receptor.
 FT NON_TER 531
 SQ SEQUENCE 531 AA; 54419 MW; FA0D3AA60AC20DAB CRC64;

Query Match 26.18; Score 142.5; DB 4; Length 531;
 Best Local Similarity 50.88; Pred. No. 2.2e-07;
 Matches 32; Conservative 3; Mismatches 12; Indels 16; Gaps 3;
 QY 18 GSGSLGKAQCAALW-LQCAAGTTI--GCGGAVACQNYRQFCRGGGGGGGGGGGGG 73
 DB 415 GSGSPSAASSSWHTLFTAEGLYPCGGG-----GGGGGGGGGGGGGGG 462
 QY 74 GGG 76
 DB 463 GGG 465

RESULT 6

Q9NUA2 PRELIMINARY; PRT; 539 AA.
 AC Q9NUA2;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE DUB0804.1 (ANDROGEN RECEPTOR (DIHYDROTESTOSTERONE RECEPTOR))
 (FRAGMENT).
 GN AR.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Chapman J.;
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 DE [2]
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Magklara A., Diamandis E.P.;
 RT "Poly-Q and poly-G repeats in the androgen receptor of the breast

cancer cell line T-47D.";
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AL049564; CAB87955.1; -;
 DR EMBL: AF321915; AAK09424.1; -;
 DR InterPro: IPR001103; Androgen_recep.
 DR Pfam: PF02166; Androgen_recep; 1.
 KW Receptor.
 FT NON_TER 539
 SQ SEQUENCE 539 AA; 55444 MW; AB493953B89D869F CRC64;
 Query Match 26.18; Score 142.5; DB 4; Length 539;
 Best Local Similarity 50.88; Pred. No. 2.2e-07;
 Matches 32; Conservative 3; Mismatches 12; Indels 16; Gaps 3;
 QY 18 GSGSLGKAQCAALW-LQCAAGTTI--GCGGAVACQNYRQFCRGGGGGGGGGGGGG 73
 DB 423 GSGSPSAASSSWHTLFTAEGLYPCGGG-----GGGGGGGGGGGGGGG 470
 QY 74 GGG 76
 DB 471 GGG 473

RESULT 7

Q9BZG6 PRELIMINARY; PRT; 542 AA.
 AC Q9BZG6;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)
 DE ANDROGEN RECEPTOR (FRAGMENT).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Magklara A., Diamandis E.P.;
 RT "Poly-Q and poly-G repeats in the androgen receptor of the breast
 cancer cell line BT-474";
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF321916; AAK09425.1; -;
 DR InterPro: IPR001103; Androgen_recep.
 DR Pfam: PF02166; Androgen_recep; 1.
 KW Receptor.
 FT NON_TER 542
 SQ SEQUENCE 542 AA; 55828 MW; C363EF841CAF7739 CRC64;

Query Match 26.18; Score 142.5; DB 4; Length 542;
 Best Local Similarity 50.88; Pred. No. 2.2e-07;
 Matches 32; Conservative 3; Mismatches 12; Indels 16; Gaps 3;
 QY 18 GSGSLGKAQCAALW-LQCAAGTTI--GCGGAVACQNYRQFCRGGGGGGGGGGGGG 73
 DB 426 GSGSPSAASSSWHTLFTAEGLYPCGGG-----GGGGGGGGGGGGGGG 473
 QY 74 GGG 76
 DB 474 GGG 476

RESULT 8

Q9BZG7 PRELIMINARY; PRT; 544 AA.
 AC Q9BZG7;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)
 DE ANDROGEN RECEPTOR (FRAGMENT).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;

```

RN [1]
RP SEQUENCE FROM N.A.
RT Magklara A., Diamandis E.P.;
RT "Poly-Q and poly-G repeats in the androgen receptor of the breast
RT cancer cell line 28-75-1."
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF321914; AAK09423.1; -
DR InterPro; IPR001103; Androgen_recep.
DR Pfam; PF02166; Androgen_recep; 1.
KW Receptor.
FT NON_TER 544 544
SQ SEQUENCE 544 AA; 56013 MW; C97133EB6C922E74 CRC64;

Query Match 26.1%; Score 142.5; DB 4; Length 544;
Best Local Similarity 50.8%; Pred. No. 2.2e-07;
Matches 32; Conservative 3; Mismatches 13; Indels 15; Gaps 3;

QY 18 GSGLGRAQCAALW--LQCAAGGTI--GCGGGAACVACNYRQFCRGGGGGGGGGGGGG 73
DB 427 GSGSPSAASSSHWTLFTAEGLQVPCGGG-----GGGGGGGGGGGGGGG 475
QY 74 GGG 76
DB 476 GGG 478

RESULT 9
O65514 PRELIMINARY; PRT; 221 AA.
AC O65514;
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE PUTATIVE GLYCINE-RICH CELL WALL PROTEIN.
GN F23E13.120 OR AT4G36230.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,
RA Heisel J., Jesse T., Heijnen L., Vos P., Mewes H.W., Mayer K.,
RA Schueller C.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RA Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,
RA Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL022141; CAAL1829.1; -
DR EMBL; AL161589; CAB80294.1; -
SQ SEQUENCE 221 AA; 20344 MW; DE8DAE2E2C57A9F8 CRC64;

Query Match 25.9%; Score 141; DB 10; Length 221;
Best Local Similarity 65.1%; Pred. No. 1.2e-07;
Matches 28; Conservative 1; Mismatches 14; Indels 0; Gaps 0;

QY 34 ASGGTTGCGGAVACVACNYRQFCRGGGGGGGGGGGGGGGGGGGGG 76
DB 130 SGGGGGGGGGGGGGGGSGRGGGGGGGGGGGGGGGGGGGGGGG 172

RESULT 10
Q38777 PRELIMINARY; PRT; 318 AA.
AC Q38777;
DT 01-NOV-1996 (TREMBlrel. 01, Created)

```

```

DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE CHITINASE PRECURSOR (FRAGMENT).
GN CHITINASE.
OS Allium sativum (Garlic).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Asparagales; Alliaceae;
OC Allium.
OX NCBI_TaxID=4682;
RN [1]
RP SEQUENCE FROM N.A.
RA van Damme E.J.M., Willens P., Peumans W.;
RT "Isolation and characterization of two different chitinase cDNA clones
RT from garlic (Allium sativum L.) shoots."
RL Submitted (MAY-1992) to the EMBL/GenBank/DBJ databases.
DR EMBL; M94105; AAA32641.1; -
DR HSP; P23951; 2BAA.
DR InterPro; IPR001002; Chitin_bind.
DR InterPro; IPR000726; Glyco_hydro_19.
DR Pfam; PF00187; chitin_binding; 1.
DR Pfam; PF00182; Glyco_hydro_19; 1.
DR PRINTS; PR00451; CHITINBINDNG.
DR ProDom; PD000574; Glyco_hydro_19; 1.
DR ProDom; PD000609; Chitin_bind; 1.
DR SMART; SM00270; ChtBD1; 1.
DR PROSITE; PS00773; CHITINASE_19_1; 1.
DR PROSITE; PS00774; CHITINASE_19_2; 1.
DR PROSITE; PS00026; CHITIN_BINDING; 1.
KW Chitin-binding; Signal.
FT NON_TER 1 1
FT SIGNAL <1 17
FT CHAIN 18 318
FT CHAIN 18 318
SQ SEQUENCE 318 AA; 33175 MW; 097C3DBD72468D4 CRC64;

Query Match 25.9%; Score 141; DB 10; Length 318;
Best Local Similarity 46.2%; Pred. No. 1.8e-07;
Matches 37; Conservative 5; Mismatches 22; Indels 16; Gaps 5;

QY 4 LPKEVKLELENKQSGIGKGAQCAALWLCQASGGTIG-----CGGGAACVACNYRQFCRG 58
DB 11 LFK-----NSYAQCGSQAGGALGALGALGALGALGALGALGALGALGALGALG 59
QY 59 GGGGGGGGGGGGGGGGGGGGGG 78
DB 60 GGGGGGGGGGGGGGGGGGGG 79

RESULT 11
Q92NU7 PRELIMINARY; PRT; 233 AA.
AC Q92NU7;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE PUTATIVE GLYCINE-RICH PROTEIN.
GN SMC04357.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-1021;
RX MEDLINE=21368234; PubMed=11474104;
RA Galibert F., Finan T.M., Long S.R., Puehler A., Abola P., Ampe F.,
RA Garibay-Hubler F., Barnett M.J., Becker A., Boistard P., Bothe G.,
RA Bounry M., Bowser L., Buhrmester J., Cadieu E., Capela D., Chain P.,
RA Cowie A., Davis R.W., Dreano S., Federspiel N.A., Fisher R.F.,
RA Gloux S., Godrie T., Goffeau A., Golding B., Gouzy J., Gurjal M.,
RA Hernandez-Lucas I., Hong A., Huizar L., Hyman K.W., Jones T., Kahn D.,
RA Kahn M.L., Kalman S., Keating D.H., Kiss E., Komp C., Lelaure V.,
RA Masny D., Palm C., Peck M.C., Pohl T.M., Portetelle D., Purnelle B.,
RA Ransperger U., Surzycki R., Thebault P., Vandenbol M.,

```

RA Vorhoelter F.J., Weidner S., Wells D.H., Wong K., Yeh K.-C., Batut J.;
RT "The composite genome of the legume symbiont *Sinorhizobium meliloti*."
RL Science 293:668-672(2001).
DR EMBL; AL591789; CAC46652.1; -.
KW Complete proteome.
SQ SEQUENCE 233 AA; 22513 MW; ED22171714E58E55 CRC64;

Query Match 25.6%; Score 139.5; DB 16; Length 233;
Best Local Similarity 48.3%; Pred. No. 1.9e-07;
Matches 29; Conservative 2; Mismatches 10; Indels 19; Gaps 2;

QY 17 KSGSLGKAQCAALWLCASGTTICGGGAVACQNYRQFCRGGGGGGGGGGGGGGG 76
DB 98 RGGGKGGG-----KGGGGVGGGG-----GGGGGGGGGGGGGGGGGG 138

RESULT 12
ID Q18880 PRELIMINARY; PRT; 393 AA.
AC Q18880;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE C56A3.1 PROTEIN.
GN C56A3.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RX [1]
RP SEQUENCE FROM N.A.
RA Sims M.A.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode *C. elegans*: A platform for
investigating biology."
RL Science 282:2012-2018(1998).
DR EMBL; 277655; CAB01137.1; -.
SQ SEQUENCE 393 AA; 37490 MW; 2843994C2FDE0F0 CRC64;

Query Match 25.6%; Score 139.5; DB 5; Length 393;
Best Local Similarity 50.8%; Pred. No. 3.3e-07;
Matches 30; Conservative 0; Mismatches 16; Indels 13; Gaps 2;

QY 18 GSGLGKAQCAALWLCASGTTICGGGAVACQNYRQFCRGGGGGGGGGGGGGGG 76
DB 87 GGGCGGGGGGGGGGGGGG--GCGGGG-----GGGGGGGGGGGGGGGGGG 132

RESULT 13
ID O65330 PRELIMINARY; PRT; 335 AA.
AC O65330;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE ACIDIC CHITINASE (EC 3.2.1.14).
OS Elaeagnus umbellata.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Rosales; Elaeagnaceae; Elaeagnus.
OX NCBI_TaxID=43233;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=ROOT NODULE;
RA Kim H.-B., An C.-S.;
RT "Isolation and characterization of cDNA clone encoding acidic
chitinase from the root nodule of *Elaeagnus umbellata*."
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF061805; AAC16010.1; -.

DR HSP; P23951; 2BAA.
DR InterPro; IPR001002; Chitin_bind.
DR InterPro; IPR000726; Glyco_hydro_19.
DR Pfam; PF00187; chitin_binding; 1.
DR Pfam; PF00182; Glyco_hydro_19; 1.
DR PRINTS; PR00451; CHITINBINDNG.
DR ProDom; PD000574; Glyco_hydro_19; 1.
DR SMART; SM00270; ChEBD1; 1.
DR PROSITE; PS00773; CHITINASE_19_1; 1.
DR PROSITE; PS00774; CHITINASE_19_2; 1.
DR PROSITE; PS00026; CHITIN_BINDING; 1.
KW Chitin-binding; Glycosidase; Hydrolase.
SQ SEQUENCE 335 AA; 35737 MW; 1CC09B8767A18E0 CRC64;

Query Match 25.5%; Score 139; DB 10; Length 335;
Best Local Similarity 51.6%; Pred. No. 3.2e-07;
Matches 33; Conservative 3; Mismatches 16; Indels 12; Gaps 4;

QY 16 QKSGSLGKAQCAALWLCASGTTIG-----CGGAVACQNYRQFCRGGGGGGGGGGGGG 70
DB 24 QCGQQLGGALCSG-GLCCSOMGYCGTDPYCGDG---CQSQ---CDGGGGGGGGGGGGG 76
QY 71 GGGG 74
DB 77 GGGG 80

RESULT 14
ID Q43522 PRELIMINARY; PRT; 207 AA.
AC Q43522;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE TFM5 PROTEIN.
GN TFM5.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. UC82B; TISSUE=FRUIT;
RX MEDLINE=97201476; PubMed=9049262;
RA Santino C.G., Stanford G.L., Conner T.W.;
RT "Developmental and transgenic analysis of two tomato fruit enhanced
genes."
RL Plant Mol. Biol. 33:405-416(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. UC82B; TISSUE=FRUIT;
RA Connor T.W.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; X95262; CAA64559.1; -.
DR HSP; P24337; 1HVP.
DR InterPro; IPR003612; AAI.
DR InterPro; IPR000528; Plant_LTP.
DR Pfam; PF00279; LTP; 1.
DR SMART; SM00499; AAI; 1.
SQ SEQUENCE 207 AA; 18265 MW; A50FE71F8ED4C4D9 CRC64;

Query Match 25.3%; Score 138; DB 10; Length 207;
Best Local Similarity 50.8%; Pred. No. 2.4e-07;
Matches 30; Conservative 1; Mismatches 6; Indels 22; Gaps 2;

QY 18 GSGLGKAQCAALWLCASGTTICGGGAVACQNYRQFCRGGGGGGGGGGGGGGG 76
DB 85 GSGSG-----GGSGTGGGGG-----GGGGGGGGGGGGGGGGGG 121

RESULT 15
Q9GRW7

```

ID Q9GRW7 PRELIMINARY; PRT; 697 AA.
AC Q9GRW7;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE DE NONA PROTEIN.
GN NONA.
OS Drosophila virilis (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7244;
RN [1]
RP SEQUENCE FROM N.A.
RA Campesan S., Chalmers D., Sandrelli F., Megighian A., Peixoto A.A.,
RA Costa R., Kyriacou C.P.;
RT "Comparative analysis of the nonA region in Drosophila identifies
RT highly diverged 5' gene that may constrain nonA promoter evolution.";
RL Submitted (Sep-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AJ298998; CAC10058.1; -.
DR HSP; F11940; IGVJ.
DR FlyBase; FBgn0042729; Dvir\nonA.
DR InterPro; IPR000504; RRM.
DR Pfam; PF00076; rrm; 2.
DR SMART; SM00360; RRM; 2.
DR PROSITE; PS50102; RRM; 2.
SQ SEQUENCE 697 AA; 75337 MW; 516BD031DA5FABFB CRC64;

Query Match 25.1%; Score 137; DB 5; Length 697;
Best Local Similarity 57.1%; Pred. No. 1.1e-06;
Matches 28; Conservative 1; Mismatches 8; Indels 12; Gaps 1;

QY 34 ASGGTIGCGGAVACQNYRCRCGGGGGGGGGGGGGGGMSKFD 82
Db 194 ARGSGGGGGG-----GGGGGGGGGGGGGGGDRDNP 230

```

Search completed: October 24, 2002, 09:29:12
Job time : 35.1168 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 09:23:41 ; Search time 14.3139 Seconds
(without alignments)
287.115 Million cell updates/sec

Title: US-09-894-030-3_COPY_1_37

Perfect score: 193
Sequence: 1 MEKLFKEVLEENKQSGSLGRAQCAALWQCASGG 37

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues
Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802 : *
1: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT : *
2: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT : *
3: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT : *
4: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT : *
5: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT : *
6: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT : *
7: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT : *
8: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT : *
9: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT : *
10: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT : *
11: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT : *
12: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT : *
13: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT : *
14: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT : *
15: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT : *
16: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT : *
17: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT : *
18: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT : *
19: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT : *
20: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT : *
21: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT : *
22: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT : *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	193	100.0	56	20	AAV03205
2	100	51.8	37	20	AAV03207
3	96	49.7	37	20	AAV03218
4	58	30.1	408	21	AAV16276
5	58	30.1	464	21	AAV16275
6	58	30.1	464	22	AAV16274
7	58	30.1	464	22	AAV20017
8	58	30.1	506	21	AAV16274
9	56	29.0	200	21	AAV58269
10	56	29.0	200	21	AAV60923
11	56	29.0	202	21	AAV58266

12	56	29.0	202	21	AAV60922	Arabidopsis thalia
13	56	29.0	215	21	AAV60921	Arabidopsis thalia
14	56	29.0	216	20	AAV58267	Arabidopsis thalia
15	55.5	28.8	70	21	AAV24125	Conopeptide propep
16	55.5	28.8	84	22	AAU05933	Cone snail O-supe
17	55	28.5	103	21	AAV02572	Human secreted pro
18	55	28.5	740	22	ABG28146	Novel human diagno
19	54.5	28.2	839	20	AAV30152	A human vanilloid
20	54.5	28.2	839	20	AAV30153	A partial human va
21	54.5	28.2	839	20	AAV30155	A human vanilloid
22	54.5	28.2	839	20	AAV06558	Human capsaisin re
23	54.5	28.2	839	21	AAV32127	Human vanilloid re
24	54.5	28.2	839	21	AAV96478	Human vanilloid re
25	54.5	28.2	839	21	AAV97357	Human VR-1 protein
26	54.5	28.2	839	22	AAE01229	Human vanilloid re
27	53.5	27.7	82	22	AAE05937	Cone snail O-supe
28	53.5	27.7	907	22	ABG15137	Thielavia terrestr
29	53	27.5	297	18	AAW20066	Fragment of cystat
30	53	27.5	467	20	AAW81021	Novel human diagno
31	53	27.5	905	22	ABG29396	Fragment of cystat
32	52	26.9	223	20	AAW81019	Gene 24 human secr
33	51.5	26.7	289	22	AAV64964	Novel human diagno
34	51.5	26.7	378	22	ABG18134	Novel human diagno
35	51.5	26.7	644	22	ABG18134	H. pylori derived
36	50.5	26.2	63	18	AAW20382	Human transcriptio
37	50.5	26.2	765	22	ABW50196	Human polypeptide
38	50.5	26.2	786	22	AAW41413	Human protein sequ
39	50.5	26.2	842	22	AAV55612	Human polypeptide
40	50.5	26.2	898	22	AAV39627	Novel human secr
41	50	25.9	221	22	AAU30375	Human polypeptide
42	50	25.9	374	22	AAV39000	Human polypeptide
43	50	25.9	375	20	AAV32885	Soybean flavanone-
44	50	25.9	481	22	AAW40786	Human polypeptide
45	50	25.9	482	21	AAV23631	Human secreted pro

ALIGNMENTS

RESULT 1
AAV03205
ID AAV03205 standard; Protein; 56 AA.
XX
AC AAV03205;
XX
DT 03-AUG-1999 (first entry)
XX
DE Amino acid sequence of sunA protein.
XX
KW Sublancin 168; antimicrobial; food preservative; pro-sublancin 168;
KW Gram-positive bacteria; pre-sublancin 168; sunA; sunT.
XX
OS Bacillus subtilis.
XX
PN WO9903352-A1.
XX
PD 28-JAN-1999.
XX
PF 17-JUL-1998; 98WO-US14547.
XX
PR 18-JUL-1997; 97US-0053035.
XX
(UYMA-) UNIV MARYLAND BALTIMORE.
XX
PI Hansen JN;
XX
DR WPI; 1999-131752/11.
DR N-PSDB; AAX28631.
XX
PT New antimicrobial peptide, sublancin 168, from Bacillus subtilis -
PT used for, e.g. treatment of infections caused by Gram negative
PT bacteria and as food preservative
XX

PS Claim 6; Fig 3; 71pp; English.

XX This is the amino acid sequence of the sunA protein encoded by the
CC prosublancin 168 nucleotide sequence used in the method of the
CC invention. The peptide designated sublancin 168, is an antimicrobial
CC useful for treating infections and preserving food against spoilage
CC bacteria, particularly Gram-positive bacteria. Pro-sublancin 168 and
CC pre-sublancin 168, are the precursors of sublancin 168. Sublancin
CC 168 is very stable at low pH and can be autoclaved without damage.
CC It does not decompose after 2 years in aqueous solution of about
CC neutral pH.
XX
SQ Sequence 56 AA;

Query Match 100.0%; Score 193; DB 20; Length 56;

Best Local Similarity 100.0%; Pred. No. 1.9e-18;

Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEKLFKEVLEENKSGSLGKACQALWLCASGG 37

DB 1 MEKLFKEVLEENKSGSLGKACQALWLCASGG 37

RESULT 2

AAAY03207

ID AAY03207 standard; Protein; 37 AA.

XX AC AAY03207;

XX DT 03-AUG-1999 (first entry)

XX DE Amino acid sequence of prosublancin 168.

XX KW Sublancin 168; antimicrobial; food preservative; pro-sublancin 168;

XX KW Gram-positive bacteria; pre-sublancin 168.

XX OS Bacillus subtilis.

XX PN WO9903352-A1.

XX PD 28-JAN-1999.

XX PF 17-JUL-1998; 98WO-US14547.

XX PR 18-JUL-1997; 97US-0053035.

XX PA (UYMA-) UNIV MARYLAND BALTIMORE.

XX PI Hansen JN;

XX PS WPI; 1999-131752/11.

XX CC This is the amino acid sequence of prosublancin 168 used in the

XX CC method of the invention. The peptide designated sublancin 168, is

XX CC an antimicrobial useful for treating infections and preserving food

XX CC against spoilage bacteria, particularly Gram-positive bacteria.

XX CC Pro-sublancin 168 and pre-sublancin 168, are the precursors of

XX CC sublancin 168. Sublancin 168 is very stable at low pH and can be

XX CC autoclaved without damage. It does not decompose after 2 years in

XX CC aqueous solution of about neutral pH.

XX SQ Sequence 37 AA;

Query Match 51.8%; Score 100; DB 20; Length 37;

Best Local Similarity 100.0%; Pred. No. 3.1e-06;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 GLGKAQCAALWLCASGG 37

DB 1 GLGKAQCAALWLCASGG 18

RESULT 3

AAAY03218

ID AAY03218 standard; Protein; 37 AA.

XX AC AAY03218;

XX DT 03-AUG-1999 (first entry)

XX DE Amino acid sequence of sublancin 168.

XX KW Sublancin 168; antimicrobial; food preservative; pro-sublancin 168;

XX KW Gram-positive bacteria; Pro-sublancin 168; pre-sublancin 168.

XX OS Bacillus subtilis.

XX FH Key Location/Qualifiers

XX FT Disulfide-bond 7..36

XX FT Disulfide-bond 14..29

XX FT Modified-site 16

XX FT Modified-site 19

XX FT Modified-site 19

XX FT Cross-links 19..22

XX FT Cross-links 19..22

XX FT Cross-links 19..22

XX FT Cross-links 19..22

XX FT Cross-links 19..22

XX FT Cross-links 19..22

XX FT Cross-links 19..22

XX FT Cross-links 19..22

XX FT Cross-links 19..22

XX FT Cross-links 19..22

XX FT Cross-links 19..22

XX FT Cross-links 19..22

XX FT Cross-links 19..22

XX FT Cross-links 19..22

XX FT Cross-links 19..22

XX FT Cross-links 19..22

XX FT Cross-links 19..22

XX FT Cross-links 19..22

XX FT Cross-links 19..22

XX FT Cross-links 19..22

XX FT Cross-links 19..22

XX FT Cross-links 19..22

XX FT Cross-links 19..22

XX FT Cross-links 19..22

XX FT Cross-links 19..22

XX FT Cross-links 19..22

XX FT Cross-links 19..22

XX FT Cross-links 19..22

XX FT Cross-links 19..22

XX FT Cross-links 19..22

XX FT Cross-links 19..22

XX FT Cross-links 19..22

XX FT Cross-links 19..22

XX FT Cross-links 19..22

XX FT Cross-links 19..22

XX FT Cross-links 19..22

XX FT Cross-links 19..22

XX FT Cross-links 19..22

XX FT Cross-links 19..22

XX FT Cross-links 19..22

XX FT Cross-links 19..22

XX FT Cross-links 19..22

XX FT Cross-links 19..22

XX FT Cross-links 19..22

XX FT Cross-links 19..22

XX FT Cross-links 19..22

XX FT Cross-links 19..22

XX FT Cross-links 19..22

XX FT Cross-links 19..22

XX FT Cross-links 19..22

XX FT Cross-links 19..22

XX FT Cross-links 19..22

XX FT Cross-links 19..22

XX FT Cross-links 19..22

PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154775.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155653.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158233.
PR 12-OCT-1999; 99US-0158363.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159323.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 30.1%; Score 58; DB 21; Length 408;
Best Local Similarity 48.6%; Pred. No. 15;
Matches 17; Conservative 2; Mismatches 10; Indels 6; Gaps 3;
QY 2 EKLEKEVLEENKGGSLGKACAAALWLOCASG 36
||| ||| :|: |||||
Db 240 EKLAKEVYF--LQNSEGSGLAPFDC--WL-CLRG 268

RESULT 5
AAG16275
ID AAG16275 standard; Protein; 464 AA.
XX
AC AAG16275;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 16856.
XX

KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135153.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.

PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140699.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.

PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154018.
PR 20-SEP-1999; 99US-0154039.
PR 22-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 28-SEP-1999; 99US-0156596.
PR 29-SEP-1999; 99US-0157117.
PR 04-OCT-1999; 99US-0157753.
PR 05-OCT-1999; 99US-0157865.
PR 06-OCT-1999; 99US-0158029.
PR 07-OCT-1999; 99US-0158232.
PR 08-OCT-1999; 99US-0158369.
PR 12-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 30.1%; Score 58; DB 21; Length 464;
Best Local Similarity 48.6%; Pred. No. 18;
Matches 17; Conservative 2; Mismatches 10; Indels 6; Gaps 3;

QY 2 EKLFKEVKLEELNKGSGLGKQAALWLOQASG 36
||| ||| | : | ||| | | | |
Db 296 EKLAKEVYF--LQNSESGSLAPFDC---NL-CLRG 324

RESULT 6
AAB47604
ID AAB47604 standard; Protein; 464 AA.
XX
AC AAB47604;

XX 07-JAN-2002 (first entry)
XX Cystathionine beta lyase, CBL.
XX Herbicide; porphobilinogen deaminase; crop; CBL; ENR-A; UROD;
KW enoyl-acyl carrier protein reductase; cystathionine beta lyase;
KW coproporphyrinogen decarboxylase; PBGD; porphobilinogen deaminase;
XX CPPO; coproporphyrinogen oxidase.
XX Arabidopsis thaliana.
XX

PN US6294345-B1.
 XX 25-SEP-2001.
 XX 05-JUN-2000; 2000US-0586719.
 XX 27-JUL-1999; 99US-198218P.
 PR 20-AUG-1999; 99US-240929P.
 PR 22-NOV-1999; 99US-228810P.
 PR 27-MAR-2000; 2000US-287572P.
 PR 22-MAY-2000; 2000US-287571P.
 XX (SYGN) SYNGENTA PARTICIPATIONS AG.
 XX Levin JZ, Bauer MW, Zheng F;
 XX WPI; 2001-638012/73.
 DR N-PSDB; AAH43576.
 XX Identifying herbicidal compounds for suppressing growth of undesirable
 PT vegetation in fields where important crops are grown, comprises
 PT combining a plant PBGD polypeptide with a test compound -
 XX Example 21; Column 55-58; 47pp; English.
 XX The sequences given in AAB4703-07 are proteins which may be used
 CC to identify herbicidal compounds, by combining one of these plant
 CC polypeptides, which have porphobilinogen deaminase activity, with a
 CC test compound under conditions allowing binding or inhibition.
 CC The method may be used for identifying herbicidal compounds, which can
 CC be used to suppress the growth of undesirable vegetation in fields where
 CC important crops are grown. The cDNA sequences encoding these proteins
 CC may be used to confer herbicide tolerance to plants, plant cells or
 CC tissues.
 XX Sequence 464 AA;
 SQ Query Match 30.1%; Score 58; DB 22; Length 464;
 Best Local Similarity 48.6%; Pred. No. 18;
 Matches 17; Conservative 2; Mismatches 10; Indels 6; Gaps 3;
 QY 2 EKLFEVKLEENKSGSLGKGAACAAALWLOCASG 36
 ||| ||| | | | | | | | | | | | | | | | |
 Db 296 EKLAKEVYF--LQNSEGSLAPFDC---WL-CLRG 324
 RESULT 7
 AAB20017
 ID AAB20017 standard; Protein; 464 AA.
 XX AC AAB20017;
 XX DT 28-MAR-2001 (first entry)
 XX DE Arabidopsis cystathionine beta lyase.
 XX KW Cystathionine beta lyase; CBL; herbicide tolerance;
 XX KW transgenic plant.
 XX OS Arabidopsis thaliana.
 XX PN WO200077185-A2.
 XX PD 21-DEC-2000.
 XX PF 13-JUN-2000; 2000WO-EP05432.
 XX PR 15-JUN-1999; 99US-0333366.
 PR 27-JUL-1999; 99US-0361879.
 PR 20-AUG-1999; 99US-0378313.
 PR 22-NOV-1999; 99US-0444117.
 PR 01-DEC-1999; 99US-0452671.
 XX

PA (NOVS) NOVARTIS AG.
 PA (NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.
 PI Levin JZ, Bauer MW, Zheng F;
 XX WPI; 2001-041382/05.
 DR N-PSDB; AAA89292.
 XX New Arabidopsis thaliana DNA molecules for identifying compounds that
 PT are inhibitors of Arabidopsis enzymes encoded by the DNA, which can be
 PT used as herbicides -
 XX Disclosure; Page 74-75; 93pp; English.
 XX The present sequence is that of Arabidopsis thaliana
 CC cystathionine beta lyase (CBL, EC-4.4.1.8), an enzyme that
 CC catalyses the conversion of cystathionine to homocysteine. The
 CC enzyme is essential for normal plant growth and development. The
 CC invention provides ENR-A, CBL, UROD, PBGD and CPPO genes (see
 CC AAA89291-95), methods for the recombinant production of the encoded
 CC enzymes (see AAB20016-20) in heterologous hosts, methods for
 CC screening chemicals for herbicidal activity using these recombinant
 CC enzymes, and methods for using these herbicides to suppress growth
 CC of undesired vegetation. Plants, plant tissue, seeds and cells
 CC tolerant to herbicides that inhibit ENR-A, CBL, UROD, PBGD or CPPO
 CC are obtained by altering the corresponding enzyme activity either by
 CC increasing expression of wild-type herbicide-sensitive enzymes or
 CC by expressing modified herbicide tolerant enzymes. The herbicide
 CC tolerant enzymes can also be used as selectable markers.
 XX Sequence 464 AA;
 SQ Query Match 30.1%; Score 58; DB 22; Length 464;
 Best Local Similarity 48.6%; Pred. No. 18;
 Matches 17; Conservative 2; Mismatches 10; Indels 6; Gaps 3;
 QY 2 EKLFEVKLEENKSGSLGKGAACAAALWLOCASG 36
 ||| ||| | | | | | | | | | | | | | | | |
 Db 296 EKLAKEVYF--LQNSEGSLAPFDC---WL-CLRG 324
 RESULT 8
 AAG16274
 ID AAG16274 standard; Protein; 506 AA.
 XX AC AAG16274;
 XX DT 17-OCT-2000 (first entry)
 XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 16855.
 XX KW Protein identification; signal transduction pathway; metabolic pathway;
 XX KW hybridisation assay; genetic mapping; gene expression control; promoter;
 XX KW termination sequence.
 XX OS Arabidopsis thaliana.
 XX PN EP1033405-A2.
 XX PD 06-SEP-2000.
 XX PF 25-FEB-2000; 2000EP-0301439.
 XX PR 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126264.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 XX

PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140891.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-APR-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149176.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 25-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.

```
PR 14-OCT-1999; 9905-0159330.
PR 14-OCT-1999; 9905-0159331.
PR 14-OCT-1999; 9905-0159637.
PR 14-OCT-1999; 9905-0159638.
PR 18-OCT-1999; 9905-0159584.
PR 21-OCT-1999; 9905-0160741.
PR 21-OCT-1999; 9905-0160767.
PR 21-OCT-1999; 9905-0160768.
PR 21-OCT-1999; 9905-0160770.
PR 21-OCT-1999; 9905-0160814.
PR 21-OCT-1999; 9905-0160815.
PR 22-OCT-1999; 9905-0160980.
PR 22-OCT-1999; 9905-0160981.
PR 22-OCT-1999; 9905-0160989.
PR 23-OCT-1999; 9905-0161404.
PR 23-OCT-1999; 9905-0161405.
PR 25-OCT-1999; 9905-0161406.
PR 26-OCT-1999; 9905-0161359.
PR 26-OCT-1999; 9905-0161360.
PR 26-OCT-1999; 9905-0161361.
PR 28-OCT-1999; 9905-0161920.
PR 28-OCT-1999; 9905-0161992.
PR 28-OCT-1999; 9905-0161993.
PR 29-OCT-1999; 9905-0162142.

Query Match 30.1%; Score 58; DB 21; Length 506;
Best Local Similarity 48.6%; Pred. No. 19;
Matches 17; Conservative 2; Mismatches 10; Indels 6; Gaps 3;

QY 2 EKLFKEVLEENKNGSGGLGKAQAALMLQASG 36
   ||| ||| | : | ||| | | | | |
DB 338 EKLAKEVYF--LQNSEGSLAPFDC---WL-CLRG 366

RESULT 9
AAG58269
ID AAG58269 standard; Protein; 200 AA.
XX
AC AAG58269;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 75196.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 9905-0121825.
PR 05-MAR-1999; 9905-0123180.
PR 09-MAR-1999; 9905-0123548.
PR 23-MAR-1999; 9905-0125788.
PR 25-MAR-1999; 9905-0126264.
PR 29-MAR-1999; 9905-0126785.
PR 01-APR-1999; 9905-0127462.
PR 06-APR-1999; 9905-0128234.
PR 08-APR-1999; 9905-0128714.
PR 16-APR-1999; 9905-0128845.
PR 19-APR-1999; 9905-0130077.
PR 21-APR-1999; 9905-0130449.
PR 23-APR-1999; 9905-0130510.
PR 23-APR-1999; 9905-0130891.
PR 28-APR-1999; 9905-0131449.
PR 30-APR-1999; 9905-0132048.
PR 30-APR-1999; 9905-0132407.

PR 04-MAY-1999; 9905-0132484.
PR 05-MAY-1999; 9905-0132485.
PR 06-MAY-1999; 9905-0132486.
PR 06-MAY-1999; 9905-0132487.
PR 07-MAY-1999; 9905-0132863.
PR 11-MAY-1999; 9905-0134256.
PR 14-MAY-1999; 9905-0134218.
PR 14-MAY-1999; 9905-0134219.
PR 14-MAY-1999; 9905-0134221.
PR 14-MAY-1999; 9905-0134370.
PR 18-MAY-1999; 9905-0134768.
PR 19-MAY-1999; 9905-0134941.
PR 20-MAY-1999; 9905-0135124.
PR 21-MAY-1999; 9905-0135353.
PR 24-MAY-1999; 9905-0135629.
PR 25-MAY-1999; 9905-0136021.
PR 27-MAY-1999; 9905-0136392.
PR 28-MAY-1999; 9905-0136782.
PR 01-JUN-1999; 9905-0137222.
PR 03-JUN-1999; 9905-0137528.
PR 04-JUN-1999; 9905-0137502.
PR 07-JUN-1999; 9905-0137724.
PR 08-JUN-1999; 9905-0138094.
PR 10-JUN-1999; 9905-0138540.
PR 10-JUN-1999; 9905-0138847.
PR 14-JUN-1999; 9905-0139119.
PR 16-JUN-1999; 9905-0139452.
PR 16-JUN-1999; 9905-0139453.
PR 17-JUN-1999; 9905-0139492.
PR 18-JUN-1999; 9905-0139454.
PR 18-JUN-1999; 9905-0139455.
PR 18-JUN-1999; 9905-0139456.
PR 18-JUN-1999; 9905-0139457.
PR 18-JUN-1999; 9905-0139458.
PR 18-JUN-1999; 9905-0139459.
PR 18-JUN-1999; 9905-0139460.
PR 18-JUN-1999; 9905-0139461.
PR 18-JUN-1999; 9905-0139462.
PR 18-JUN-1999; 9905-0139463.
PR 18-JUN-1999; 9905-0139750.
PR 18-JUN-1999; 9905-0139763.
PR 21-JUN-1999; 9905-0139817.
PR 22-JUN-1999; 9905-0139899.
PR 23-JUN-1999; 9905-0140353.
PR 23-JUN-1999; 9905-0140354.
PR 24-JUN-1999; 9905-0140695.
PR 28-JUN-1999; 9905-0140823.
PR 29-JUN-1999; 9905-0140991.
PR 30-JUN-1999; 9905-0141287.
PR 01-JUL-1999; 9905-0141842.
PR 01-JUL-1999; 9905-0142154.
PR 02-JUL-1999; 9905-0142055.
PR 06-JUL-1999; 9905-0142390.
PR 08-JUL-1999; 9905-0142803.
PR 09-JUL-1999; 9905-0142920.
PR 12-JUL-1999; 9905-0142977.
PR 13-JUL-1999; 9905-0143542.
PR 14-JUL-1999; 9905-0143624.
PR 15-JUL-1999; 9905-0144005.
PR 16-JUL-1999; 9905-0144085.
PR 16-JUL-1999; 9905-0144086.
PR 19-JUL-1999; 9905-0144325.
PR 19-JUL-1999; 9905-0144331.
PR 19-JUL-1999; 9905-0144332.
PR 19-JUL-1999; 9905-0144333.
PR 19-JUL-1999; 9905-0144334.
PR 19-JUL-1999; 9905-0144335.
PR 20-JUL-1999; 9905-0144352.
PR 20-JUL-1999; 9905-0144632.
PR 20-JUL-1999; 9905-0144884.
PR 21-JUL-1999; 9905-0144814.
PR 21-JUL-1999; 9905-0145086.
PR 21-JUL-1999; 9905-0145088.
```


PR	02-AUG-1999;	99US-0146389.	PR	28-OCT-1999;	99US-0161992.	
PR	03-AUG-1999;	99US-0147038.	PR	28-OCT-1999;	99US-0161993.	
PR	04-AUG-1999;	99US-0147204.	PR	29-OCT-1999;	99US-0162142.	
PR	04-AUG-1999;	99US-0147302.				
PR	05-AUG-1999;	99US-0147192.				
PR	05-AUG-1999;	99US-0147260.				
PR	06-AUG-1999;	99US-0147303.				
PR	06-AUG-1999;	99US-0147416.				
PR	09-AUG-1999;	99US-0147493.				
PR	09-AUG-1999;	99US-0147933.				
PR	10-AUG-1999;	99US-0148171.				
PR	11-AUG-1999;	99US-0148319.				
PR	12-AUG-1999;	99US-0148341.				
PR	13-AUG-1999;	99US-0148565.				
PR	13-AUG-1999;	99US-0148684.				
PR	16-AUG-1999;	99US-0149368.				
PR	17-AUG-1999;	99US-0149175.				
PR	18-AUG-1999;	99US-0149426.				
PR	20-AUG-1999;	99US-0149722.				
PR	20-AUG-1999;	99US-0149723.				
PR	20-AUG-1999;	99US-0149923.				
PR	23-AUG-1999;	99US-0149902.				
PR	23-AUG-1999;	99US-0149930.				
PR	25-AUG-1999;	99US-0150566.				
PR	26-AUG-1999;	99US-0151438.				
PR	26-AUG-1999;	99US-0151930.				
PR	27-AUG-1999;	99US-0150884.				
PR	27-AUG-1999;	99US-0151065.				
PR	27-AUG-1999;	99US-0151066.				
PR	27-AUG-1999;	99US-0151080.				
PR	30-AUG-1999;	99US-0151303.				
PR	31-AUG-1999;	99US-0151438.				
PR	01-SEP-1999;	99US-0151930.				
PR	07-SEP-1999;	99US-0152363.				
PR	10-SEP-1999;	99US-0153070.				
PR	13-SEP-1999;	99US-0153758.				
PR	15-SEP-1999;	99US-0154018.				
PR	16-SEP-1999;	99US-0154039.				
PR	20-SEP-1999;	99US-0154779.				
PR	22-SEP-1999;	99US-0155135.				
PR	23-SEP-1999;	99US-0155186.				
PR	24-SEP-1999;	99US-0155659.				
PR	28-SEP-1999;	99US-0156458.				
PR	29-SEP-1999;	99US-0156596.				
PR	04-OCT-1999;	99US-0157117.				
PR	05-OCT-1999;	99US-0157753.				
PR	06-OCT-1999;	99US-0157865.				
PR	07-OCT-1999;	99US-0158029.				
PR	08-OCT-1999;	99US-0158232.				
PR	12-OCT-1999;	99US-0158369.				
PR	13-OCT-1999;	99US-0159293.				
PR	13-OCT-1999;	99US-0159294.				
PR	13-OCT-1999;	99US-0159295.				
PR	14-OCT-1999;	99US-0159329.				
PR	14-OCT-1999;	99US-0159330.				
PR	14-OCT-1999;	99US-0159331.				
PR	14-OCT-1999;	99US-0159637.				
PR	14-OCT-1999;	99US-0159638.				
PR	18-OCT-1999;	99US-0159584.				
PR	21-OCT-1999;	99US-0160741.				
PR	21-OCT-1999;	99US-0160767.				
PR	21-OCT-1999;	99US-0160768.				
PR	21-OCT-1999;	99US-0160770.				
PR	21-OCT-1999;	99US-0160814.				
PR	21-OCT-1999;	99US-0160815.				
PR	22-OCT-1999;	99US-0160980.				
PR	22-OCT-1999;	99US-0160981.				
PR	22-OCT-1999;	99US-0160989.				
PR	25-OCT-1999;	99US-0161404.				
PR	25-OCT-1999;	99US-0161405.				
PR	25-OCT-1999;	99US-0161406.				
PR	26-OCT-1999;	99US-0161359.				

PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140931.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.

PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 23-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 29.0%; Score 56; DB 21; Length 202;
Best Local Similarity 43.3%; Pred. No. 13;
Matches 13; Conservative 5; Mismatches 10; Indels 2; Gaps 1;

Qy 4 LFKVLELEENOKGSGKAOCAAL--WL 31
| | | : : | | | : | | | : | |
Db 130 LFKKLDVEGLSEAIAGLAAMGCAAFWL 159

RESULT 13
AAG60921
ID AAG60921 standard; Protein; 215 AA.
XX AC AAG60921;
XX DT 18-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 78964.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
PR 03-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0128645.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 18-JUN-1999; 99US-0139817.
PR 21-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 21-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147453.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.

DE	Conopeptide propeptide Mr6.3.
XX	
KW	Conopeptide; gamma-conopeptide; venom; cone snail; cation channel;
KW	epilepsy; pacemaker; heart muscle; neuronal pacemaker calcium channel.
XX	
OS	Conus marmoreus.
XX	
PN	WO930732-A1.
XX	
PD	24-JUN-1999.
XX	
PF	16-DEC-1998; 98WO-US26792.
XX	
PR	16-DEC-1997; 97US-0069706.
XX	
PA	(REGC) UNIV CALIFORNIA.
PA	(UTAH) UNIV UTAH RES FOUND.
PA	(UYVR-) UNIV VRIJE.
XX	
PI	Burlingame AL, Colledge C, Cruz LJ, Fainzilber M;
PI	Imperial J, Kits KS, Ollivera BM, Shetty R, Walker C;
XX	Walkins M;
XX	
DR	WPI; 1999-418708/35.
DR	N-PSDB; AAX88505.
XX	
PT	Gamma-carboxylated conopeptides used as, e.g. neuronal pacemaker
PT	calcium channels
XX	
PS	Claim 37; Page 52; 61pp; English.
XX	
CC	The present invention describes gamma-carboxylated conopeptides derived
CC	from cone snail venom. The gamma-conopeptides and their propeptides are
CC	useful as agonists of neuronal pacemaker calcium channels. The
CC	conopeptides are naturally available in minute amounts in the venom of
CC	cone snails and their derivatives are synthetic. The peptides modulate
CC	slow inward cation channels in vertebrates involved in syndromes of
CC	clinical relevance, such as epileptic activity in hippocampus and
CC	pacemaker potentials in heart muscle. The present sequence represents
CC	a conopeptide propeptide.
XX	
XX	Sequence 70 AA:
XX	
SO	

Search completed: October 24, 2002, 09:28:00
Job time : 16.3139 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 09:27:01 ; Search time 5.67153 Seconds
(without alignments)
159.348 Million cell updates/sec

Title: US-09-894-030-3_COPY_1_37
Perfect score: 193
Sequence: 1 MEKLFKEVLEENKQSGIGKACAAALWLOQASGG 37

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues
Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA: *
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	58	30.1	464	4	US-09-586-719-4
2	54.5	28.2	839	4	US-09-197-636-2
3	54.5	28.2	839	4	US-09-197-636-4
4	54.5	28.2	839	4	US-09-197-636-8
5	54.5	28.2	839	4	US-09-235-451-34
6	51	26.4	549	1	US-08-325-071-61
7	51	26.4	549	4	US-08-461-004A-61
8	51	26.4	620	1	US-08-325-071-65
9	51	26.4	620	4	US-08-461-004A-65
10	51	26.4	650	1	US-08-325-071-63
11	51	26.4	650	1	US-08-325-071-67
12	51	26.4	650	4	US-08-461-004A-63
13	51	26.4	650	4	US-08-461-004A-67
14	49	25.4	259	2	US-08-402-804-8
15	49	25.4	1296	3	US-08-470-260-3
16	49	25.4	1296	3	US-08-471-491-3
17	49	25.4	1296	4	US-08-466-662-3
18	48.5	25.1	1287	1	US-08-200-232-2
19	48.5	25.1	1287	5	PCT-US95-02219-2
20	48.5	25.1	1287	5	PCT-US95-02219A-2
21	48	24.9	581	2	US-08-724-394A-3
22	47	24.4	428	3	US-09-118-319-5
23	47	24.4	500	3	US-09-141-000-2
24	47	24.4	1040	2	US-08-254-989-2
25	46.5	24.1	728	4	US-08-981-392-2
26	46.5	24.1	729	3	US-08-872-855-8
27	46	23.8	215	4	US-09-220-528-104

28	46	23.8	284	5	PCT-US94-09752-4
29	46	23.8	591	1	US-08-145-995A-21
30	46	23.8	591	2	US-08-451-747-21
31	46	23.8	591	3	US-09-134-852-21
32	46	23.8	650	1	US-08-325-071-56
33	46	23.8	650	1	US-08-325-071-59
34	46	23.8	650	4	US-08-461-004A-56
35	46	23.8	650	4	US-08-461-004A-59
36	46	23.8	688	1	US-08-325-071-57
37	46	23.8	688	4	US-08-461-004A-57
38	45.5	23.6	838	4	US-09-235-451-2
39	45	23.3	82	1	US-08-319-554A-6
40	45	23.3	82	1	US-08-624-123-6
41	45	23.3	82	2	US-09-006-156-6
42	45	23.3	147	1	US-08-325-071-69
43	45	23.3	147	4	US-08-461-004A-69
44	44.5	23.1	200	4	US-09-068-740A-2
45	44.5	23.1	520	4	US-09-088-740A-3

ALIGNMENTS

RESULT 1
US-09-586-719-4
; Sequence 4, Application US/09586719
; Patent No. 6294345
; GENERAL INFORMATION:
; APPLICANT: Zheng, Feng
; APPLICANT: Levin, Joshua Z.
; APPLICANT: Bauer, Michael W.
; TITLE OF INVENTION: Herbicide Target Genes and Methods
; FILE REFERENCE: PB/5-30909A
; CURRENT APPLICATION NUMBER: US/09/586,719
; CURRENT FILING DATE: 2000-06-05
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 4
; LENGTH: 464
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-586-719-4

Query Match 30.1%; Score 58; DB 4; Length 464;
Best Local Similarity 48.6%; Pred. No. 4.4;
Matches 17; Conservative 2; Mismatches 10; Indels 6; Gaps 3;

QY 2 EKLFKEVLEENKQSGIGKACAAALWLOQASG 36
Db 296 EKLAKEVIF--LQNSEGSLAPFDC---WL-CLRG 324

RESULT 2
US-09-197-636-2
; Sequence 2, Application US/09197636
; Patent No. 6239267
; GENERAL INFORMATION:
; APPLICANT: DUCKWORTH, DAVID
; APPLICANT: HAYES, PHILIP
; APPLICANT: MEADOWS, HELEN
; APPLICANT: DAVIS, JOHN
; TITLE OF INVENTION: NOVEL COMPOUNDS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ratner & Prestia
; STREET: P.O. Box 980
; CITY: Valley Forge
; STATE: PA
; COUNTRY: US
; ZIP: 19482-0980
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible

```
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/197,636
; FILING DATE: 23-NOV-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: UK 9805137.8
; FILING DATE: 12-MAR-1998
; APPLICATION NUMBER: UK 9815791.0
; FILING DATE: 21-JUL-1998
; APPLICATION NUMBER: UK 9819278.4
; FILING DATE: 03-SEP-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Prestia, Paul F
; REGISTRATION NUMBER: 23,031
; REFERENCE/DOCKET NUMBER: GP-30075
; TELEPHONE: 601-407-0700
; TELEFAX: 610-407-0701
; TELEX: 846169
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 839 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-197-636-2

Query Match      28.2%; Score 54.5; DB 4; Length 839;
Best Local Similarity 43.2%; Pred. No. 24;
Matches 16; Conservative 3; Mismatches 11; Indels 7; Gaps 1;

QY      3 KLFKEVKLELENQKG-----SGLGKAQCAALWLQ 32
      || :||||| :||| :||| :||| :|||
Db      319 KLHPTLKLEELTNKKGMTPLAALAGTGKIGVLAIIQ 355

RESULT 3
US-09-197-636-4
; Sequence 4, Application US/09197636
; Patent No. 6239267
; GENERAL INFORMATION:
; APPLICANT: DUCKWORTH, DAVID
; APPLICANT: HAYES, PHILIP
; APPLICANT: MEADOWS, HELEN
; APPLICANT: DAVIS, JOHN
; TITLE OF INVENTION: NOVEL COMPOUNDS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ratner & Prestia
; STREET: P.O. Box 980
; CITY: Valley Forge
; STATE: PA
; COUNTRY: US
; ZIP: 19482-0980
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/197,636
; FILING DATE: 23-NOV-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: UK 9805137.8
; FILING DATE: 12-MAR-1998
; APPLICATION NUMBER: UK 9815791.0
; FILING DATE: 21-JUL-1998
; APPLICATION NUMBER: UK 9819278.4
; FILING DATE: 03-SEP-1998
```

```
; ATTORNEY/AGENT INFORMATION:
; NAME: Prestia, Paul F
; REGISTRATION NUMBER: 23,031
; REFERENCE/DOCKET NUMBER: GP-30075
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 601-407-0700
; TELEFAX: 610-407-0701
; TELEX: 846169
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 839 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-197-636-4

Query Match      28.2%; Score 54.5; DB 4; Length 839;
Best Local Similarity 43.2%; Pred. No. 24;
Matches 16; Conservative 3; Mismatches 11; Indels 7; Gaps 1;

QY      3 KLFKEVKLELENQKG-----SGLGKAQCAALWLQ 32
      || :||||| :||| :||| :||| :|||
Db      319 KLHPTLKLEELTNKKGMTPLAALAGTGKIGVLAIIQ 355

RESULT 4
US-09-197-636-8
; Sequence 8, Application US/09197636
; Patent No. 6239267
; GENERAL INFORMATION:
; APPLICANT: DUCKWORTH, DAVID
; APPLICANT: HAYES, PHILIP
; APPLICANT: MEADOWS, HELEN
; APPLICANT: DAVIS, JOHN
; TITLE OF INVENTION: NOVEL COMPOUNDS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ratner & Prestia
; STREET: P.O. Box 980
; CITY: Valley Forge
; STATE: PA
; COUNTRY: US
; ZIP: 19482-0980
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/197,636
; FILING DATE: 23-NOV-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: UK 9805137.8
; FILING DATE: 12-MAR-1998
; APPLICATION NUMBER: UK 9815791.0
; FILING DATE: 21-JUL-1998
; APPLICATION NUMBER: UK 9819278.4
; FILING DATE: 03-SEP-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Prestia, Paul F
; REGISTRATION NUMBER: 23,031
; REFERENCE/DOCKET NUMBER: GP-30075
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 601-407-0700
; TELEFAX: 610-407-0701
; TELEX: 846169
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 839 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
```

COMPUTER READABLE FORM:

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/461,004A
; FILING DATE: 04-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/325,071
; FILING DATE: 19-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/062,109
; FILING DATE: 17-MAY-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/926,368
; FILING DATE: 07-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/242,196
; FILING DATE: 06-JUL-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/AU87/00401
; FILING DATE: 27-NOV-1987
; APPLICATION NUMBER: AU PI4912
; FILING DATE: 16-OCT-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU PI2570
; FILING DATE: 19-JUN-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU PH9196
; FILING DATE: 27-NOV-1986
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 60042/152
; TELEPHONE: 202 672 5300
; TELEFAX: 202 672 5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 61:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 549 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-461-004A-61

```

```

Query Match 26.4%; Score 51; DB 4; Length 549;
Best Local Similarity 35.7%; Pred. No. 46;
Matches 10; Conservative 6; Mismatches 8; Indels 4; Gaps 1;

```

```

Qy 13 LENQKSGGLGKAQCAAL---WLOQASG 36
|:::|:::|:::|:::|:::|:::|
Db 361 LKQEAAYKQKCKVKVDNLFWFQCADG 388

```

```

RESULT 8
US-08-325-071-65
; Sequence 65, Application US/08325071
; Patent No. 5587311
; GENERAL INFORMATION:
; APPLICANT: COBON, Stewart Gary
; APPLICANT: MOORE, Joanna Terry
; APPLICANT: JOHNSON, Law Anthony York
; APPLICANT: WILLADSEN, Peter
; APPLICANT: KEMP, David Harold
; APPLICANT: SRISKANTHA, Alagacone
; APPLICANT: RIDING, George Alfred
; APPLICANT: RAND, Keith No. 5587311man
; TITLE OF INVENTION: DNA Encoding A Cell Membrane
; TITLE OF INVENTION: Glycoprotein Of A Tick Gut
; NUMBER OF SEQUENCES: 71
; CORRESPONDENCE ADDRESS:

```

```

; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/325,071
; FILING DATE: 14-OCT-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/062,109
; FILING DATE: 17-MAY-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/926,368
; FILING DATE: 07-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/242,196
; FILING DATE: 06-JUL-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/AU87/00401
; FILING DATE: 27-NOV-1987
; APPLICATION NUMBER: AU PI4912
; FILING DATE: 16-OCT-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU PI2570
; FILING DATE: 19-JUN-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU PH9196
; FILING DATE: 27-NOV-1986
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 60042/111 BIAU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 672 5300
; TELEFAX: 202 672 5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 620 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-325-071-65

```

```

Query Match 26.4%; Score 51; DB 1; Length 620;
Best Local Similarity 35.7%; Pred. No. 52;
Matches 10; Conservative 6; Mismatches 8; Indels 4; Gaps 1;

```

```

Qy 13 LENQKSGGLGKAQCAAL---WLOQASG 36
|:::|:::|:::|:::|:::|:::|
Db 412 LKQEAAYKQKCKVKVDNLFWFQCADG 439

```

```

RESULT 9
US-08-461-004A-65
; Sequence 65, Application US/08461004A
; Patent No. 6235283
; GENERAL INFORMATION:
; APPLICANT: COBON, Stewart Gary
; APPLICANT: MOORE, Joanna Terry
; APPLICANT: JOHNSON, Law Anthony York
; APPLICANT: WILLADSEN, Peter
; APPLICANT: KEMP, David Harold
; APPLICANT: SRISKANTHA, Alagacone
; APPLICANT: RIDING, George Alfred
; APPLICANT: RAND, Keith No. 6235283man
; TITLE OF INVENTION: DNA Encoding A Cell Membrane

```

```

1 APPLICANT: WILLADSEN, Peter
2 APPLICANT: KEMP, David Harold
3 APPLICANT: SRISKANTHA, Alagosome
4 APPLICANT: RIDING, George Alfred
5 APPLICANT: RAND, Keith NO. 5587311man
6 TITLE OF INVENTION: DNA Encoding A Cell Membrane
7 PROTEIN OF INVENTION: Glycoprotein Of A Tick Gut
8 NUMBER OF SEQUENCES: 71
9 CORRESPONDENCE ADDRESS:
10 ADDRESSEE: Foley & Lardner
11 STREET: 3000 K Street, N.W.
12 CITY: Washington, D.C.
13 COUNTRY: USA
14 ZIP: 20007-5109
15 COMPUTER READABLE FORM:
16 MEDIUM TYPE: Floppy disk
17 COMPUTER: IBM PC compatible
18 OPERATING SYSTEM: PC-DOS/MS-DOS
19 SOFTWARE: PatentIn Release #1.0, Version #1.25
20 CURRENT APPLICATION DATA:
21 APPLICATION NUMBER: US/08/325,071
22 FILING DATE: 14-OCT-1993
23 PRIOR APPLICATION DATA:
24 PRIOR APPLICATION NUMBER: US 08/062,109
25 FILING DATE: 17-MAY-1993
26 PRIOR APPLICATION DATA:
27 APPLICATION NUMBER: US 07/926,368
28 FILING DATE: 07-AUG-1992
29 PRIOR APPLICATION DATA:
30 APPLICATION NUMBER: 07/242,196
31 FILING DATE: 06-JUL-1988
32 PRIOR APPLICATION DATA:
33 APPLICATION NUMBER: PCT/AU87/00401
34 FILING DATE: 27-NOV-1987
35 PRIOR APPLICATION DATA:
36 APPLICATION NUMBER: AU PI4912
37 FILING DATE: 16-OCT-1987
38 PRIOR APPLICATION DATA:
39 APPLICATION NUMBER: AU PI2570
40 FILING DATE: 19-JUN-1987
41 PRIOR APPLICATION DATA:
42 APPLICATION NUMBER: AU PH9196
43 FILING DATE: 27-NOV-1986
44 ATTORNEY/AGENT INFORMATION:
45 NAME: BENT, Stephen A.
46 REGISTRATION NUMBER: 29,768
47 REFERENCE/DOCKET NUMBER: 60042/111 BIAU
48 TELECOMMUNICATION INFORMATION:
49 TELEPHONE: 202 672 5300
50 TELEFAX: 202 672 5399
51 TELEX: 904136
52 INFORMATION FOR SEQ ID NO: 63:
53 SEQUENCE CHARACTERISTICS:
54 LENGTH: 650 amino acids
55 TYPE: amino acid
56 TOPOLOGY: linear
57 MOLECULE TYPE: protein
58 OS-08-325-071-63

```

Query Match 26.4%; Score 51; DB 1; Length 650;
Best Local Similarity 35.7%; pred. No. 55;
Matches 10; Conservative 6; Mismatches 8; Indels 4; Gaps 1;

```
QY      13  LENKQSGGLGKAQCAAL---WLQCASG 36
      1:1:1: : 1: : 1: : 1:1:1:1:
Db      442  LKNQEAAYKQGNKCCKVKVDNLFWFQCADG 469
```

US-08-325-071-63
; Sequence 63, Application US/08325071
; Patent No. 5587311

RESULT 11
US-08-325-071-67
; Sequence 67, Application US/08325071
; Patent No. 5587311
; GENERAL INFORMATION:

APPLICANT: COBON, Stewart Gary
APPLICANT: MOORE, Joanna Terry
APPLICANT: JOHNSON, Law Anthony York
APPLICANT: WILLADSEN, Peter
APPLICANT: KEMP, David Harold
APPLICANT: SRISKANTHA, Alagacone
APPLICANT: RIDING, George Alfred
APPLICANT: RAND, Keith No. 5587311man
TITLE OF INVENTION: DNA Encoding A Cell Membrane
TITLE OF INVENTION: Glycoprotein Of A Tick Gut
NUMBER OF SEQUENCES: 71
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W.
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/325,071
FILING DATE: 14-OCT-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/062,109
FILING DATE: 17-MAY-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/926,368
FILING DATE: 07-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/242,196
FILING DATE: 06-JUL-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU P12570
FILING DATE: 19-JUN-1987
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PH9196
FILING DATE: 27-NOV-1986
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 60042/111 BIAU
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202 672 5300
TELEFAX: 202 672 5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 67:
SEQUENCE CHARACTERISTICS:
LENGTH: 650 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-325-071-67

Query Match 26.4%; Score 51; DB 1; Length 650;
Best Local Similarity 35.7%; Pred. No. 55;
Matches 10; Conservative 6; Mismatches 8; Indels 4; Gaps 1;
QY 13 LENQKSGSLGKAQCAAL----WLCQASG 36
Db 442 LKNOEAAAYKGNCKVKVDNLFNFQCADG 469

RESULT 12
US-08-461-004A-63

Sequence 63, Application US/08461004A
Patent No. 6235283
GENERAL INFORMATION:
APPLICANT: COBON, Stewart Gary
APPLICANT: MOORE, Joanna Terry
APPLICANT: JOHNSON, Law Anthony York
APPLICANT: WILLADSEN, Peter
APPLICANT: KEMP, David Harold
APPLICANT: SRISKANTHA, Alagacone
APPLICANT: RIDING, George Alfred
APPLICANT: RAND, Keith No. 6235283man
TITLE OF INVENTION: DNA Encoding A Cell Membrane
TITLE OF INVENTION: Glycoprotein Of A Tick Gut
NUMBER OF SEQUENCES: 71
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W.
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/461,004A
FILING DATE: 04-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/325,071
FILING DATE: 19-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/062,109
FILING DATE: 17-MAY-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/926,368
FILING DATE: 07-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/242,196
FILING DATE: 06-JUL-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU87/00401
FILING DATE: 27-NOV-1987
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU P14912
FILING DATE: 16-OCT-1987
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU P12570
FILING DATE: 19-JUN-1987
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PH9196
FILING DATE: 27-NOV-1986
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 60042/152
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202 672 5300
TELEFAX: 202 672 5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 63:
SEQUENCE CHARACTERISTICS:
LENGTH: 650 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-461-004A-63

Query Match 26.4%; Score 51; DB 4; Length 650;
Best Local Similarity 35.7%; Pred. No. 55;
Matches 10; Conservative 6; Mismatches 8; Indels 4; Gaps 1;
QY 13 LENQKSGSLGKAQCAAL----WLCQASG 36

Query Match 26.4%; Score 51; DB 4; Length 650;
Best Local Similarity 35.7%;
Matches 10; Conservative 6; Mismatches 8; Indels
Pred. No. 55;

OV 13 LENQKSGSLGKAQCAAL---WLQCSG 36

```

QY      13 LENQKSGGLGKACCAAL- ---WLQCASG 36
       |:|: : |:| : |:| |
Db      442 LKNQEAAAYKGONCKVKVDNLFWFQCADG 469

RESULT 14
US-08-402-804-8
; Sequence 8, Application US/08402804
; Patent No. 5874300
; GENERAL INFORMATION:
; APPLICANT: Blaser, Martin J.
; APPLICANT: Pei, Zhiheng
; TITLE OF INVENTION: Campylobacter Jejuni Antigens, And
; TITLE OF INVENTION: Methods For Their Production And Use
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OSTROLENK, FABER, GERB & SOFFEN
; STREET: 1180 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-8403
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/402-804

```

FILING DATE: 05/05/1994
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/114,420
 FILING DATE: 30-AUG-1993
 APPLICATION NUMBER: US 08/112,387
 FILING DATE: 27-AUG-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/986,928
 FILING DATE: 08-DEC-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/612,330
 FILING DATE: 13-NOV-1990
 ATTORNEY/AGENT INFORMATION:
 NAME: Gray III, William O.
 REGISTRATION NUMBER: 30,944
 REFERENCE/DOCKET NUMBER: P/1261-13
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 382-0700
 TELEFAX: (212) 382-0888
 TELEX: 236925
 INFORMATION FOR SEQ ID NO: 8:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 259 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-402-804-8

RESULT 15

Query Match 25.4%; Score 49; DB 2; Length 259;
Best Local Similarity 41.7%; Pred. No. 38;
Matches 10: Conservative 5: Mismatches 9: Indels

4 LFKEVKLEELNQKSGIGKAQCA 27
: ||| : ||| : ||| : ||| :
134 VLKEKYKSLADMKGANIGVAQAA 157

US-08-470-260-3
; Sequence 3, Application US/08470260
; Patent No. 6077706
; GENERAL INFORMATION:
; APPLICANT: Covacci, Antonello
; APPLICANT: Bugnoli, Massimo
; APPLICANT: Telford, John
; APPLICANT: Macchia, Giovanni
; APPLICANT: Rappuoli, Rino
; TITLE OF INVENTION: Helicobacter Pylori Proteins Useful
; TITLE OF INVENTION: for Vaccines and Diagnostics
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: California
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/470,260
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/256,848
; FILING DATE: 21-OCT-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0316.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1296 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-470-260-3

Query Match 25.4%; Score 49; DB 3; Length 1296;
Best Local Similarity 37.1%; Pred. No. 2.1e+02;
Matches 13; Conservative 6; Mismatches 12; Indels 4; Gaps 2;

QY 6 KEVKLE--ELENKQSGGLG-KAQCAALWLOCASG 36
Db 193 KNISIDNFVEINNRVSGAGRRASSTVLTQASEG 227

Search completed: October 24, 2002, 09:30:13
Job time : 6.67153 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 09:26:16 ; Search time 7.0219 Seconds
(without alignments)
506.317 Million cell updates/sec

Title: US-09-894-030-3_COPY_1_37

Perfect score: 193
Sequence: 1 MEKLFKEVKELEENQKSGLGKQAALMLQCASGG 37

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	193	100.0	56	T12783	sublancin 168 prec
2	61	31.6	860	S43846	xylanase B - rumen
3	58	30.1	464	S61429	cystathionine beta
4	56	29.0	202	D84607	hypothetical prote
5	54.5	28.2	839	JC7621	capsaicin receptor
6	53	27.5	1000	T13636	probable minor str
7	53	27.5	1006	T13331	probable tail prot
8	52	26.9	469	AE1220	ethanolamine utili
9	52	26.9	761	T05299	hypothetical prote
10	51.5	26.7	295	D90252	conserved hypothet
11	51.5	26.7	564	HM1VF6	hemagglutinin prec
12	51.5	26.7	1024	E86331	hypothetical prote
13	51.5	26.7	2584	T24158	hypothetical prote
14	51.5	26.7	2606	T24157	hypothetical prote
15	51	26.4	361	C83350	probable transcrip
16	51	26.4	533	D96777	hypothetical prote
17	51	26.4	638	F70444	threonine--trna li
18	51	26.4	770	T50308	probable translati
19	51	26.4	1738	C84507	GTP-binding protei
20	50.5	26.2	191	S54295	caspo homolog - fis
21	50.5	26.2	633	T41332	hypothetical prote
22	50	25.9	148	D71007	hypothetical prote
23	49.5	25.6	564	HM1VF5	hemagglutinin prec
24	49.5	25.6	564	HM1VF7	hemagglutinin prec
25	49.5	25.6	564	HM1VF8	hemagglutinin prec
26	49.5	25.6	564	HM1VF9	hemagglutinin prec
27	49.5	25.6	568	A46339	peptide-aspartate
28	49.5	25.6	754	BABOH	hypothetical prote
29	49.5	25.6	764	C86314	hypothetical prote

adenylate cyclase
hypothetical prote
hypothetical prote
probable ABC-type
probable pdhc prot
tyrosyl-trna synth
vacuolating cyto
two-component hyb
phage-related prot
hypothetical prote
heterogeneous nucl
modulation protein
pmt2 methyltransf
hypothetical prote
vacuolating cyto
vacuolating cyto

30 49.5 25.6 781 2 AF2516
31 49 25.4 243 2 A99995
32 49 25.4 243 2 F85840
33 49 25.4 259 2 A48518
34 49 25.4 393 2 G70549
35 49 25.4 407 2 C96978
36 49 25.4 1288 2 E71884
37 49 25.4 1627 2 AE2109
38 48.5 25.1 111 2 C69949
39 48.5 25.1 200 2 T48130
40 48.5 25.1 534 1 S66820
41 48.5 25.1 560 2 PQ0041
42 48.5 25.1 802 2 T37754
43 48.5 25.1 1122 2 T28130
44 48.5 25.1 1287 2 B53739
45 48.5 25.1 1290 2 G64630

RESULT 1
T12783
sublancin 168 precursor - Bacillus subtilis phage SPBc2
C:Species: Bacillus subtilis phage SPBc2
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 24-Sep-1999
C:Accession: T12783; H69719
R:Lazarevic, V.; Duesterhoeft, A.; Soldo, B.; Hilbert, H.; Mauel, C.; Karamata, D.
submitted to the EMBL Data Library, August 1997
A:Description: The complete nucleotide sequence of the Bacillus subtilis SPbetac2 pro
A:Reference number: Z17583
A:Accession: T12783
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-56 <LAZ>
A:Cross-references: EMBL:AF020713; NID:93025478; PID:g3025497; PIDN:AA012992.1
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.; Ehrlich, S.D.; Emmerison, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scani
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se
Akeuchi, M.; Tanakoshi, A.; Tanaka, T.; Terptrat, P.; Tognoni, A.; Tosato, V.; Uchiya
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtil
A:Reference number: A69580; MUID:98044033
A:Accession: H69719
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-56 <KUN>
A:Cross-references: GB:Z99115; GB:AL009126; NID:g2634478; PIDN:CABL4066.1; PID:ell1835
A:Experimental source: strain 168
C:Genetics: <LAI>
A:Gene: yolG
C:Genetics: <KUI>
A:Gene: sunA
C:Superfamily: unassigned lanthionine-containing peptides
C:Keywords: antibiotic; lanthionine
F:1-19/Domain: propeptide #status predicted <PRO>
F:20-56/Product: sublancin 168 #status predicted <MAT>
F:26-55/Disulfide bonds: #status experimental
F:33-48/Disulfide bonds: #status predicted
F:35/Modified site: dehydroalanine (Ser) #status experimental
F:38-41/Cross-link: (2S,3S,6R)-3-methyl-lanthionine (Thr-Cys) #status predicted

Query Match 100.0%; Score 193; DB 2; Length 56;
Best Local Similarity 100.0%; Pred. No. 3.4e-18;

Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEKLFKEVKLELENKQSGGLGKAQCAALWLQCSGG 37
 |||||
 DB 1 MEKLFKEVKLELENKQSGGLGKAQCAALWLQCSGG 37

RESULT 2

S43846
 xylanase B - rumen fungus (Neocallimastix patriciarum)
 C:Species: Neocallimastix patriciarum
 C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 28-May-1999
 C:Accession: S43846
 R:Black, G.W.; Hazlewood, G.P.; Xue, G.P.; Orpin, C.G.; Gilbert, H.J.
 Biochem. J. 299, 381-387, 1994
 A:Title: Xylanase B from Neocallimastix patriciarum contains a non-catalytic 455-residue
 A:Reference number: S43846; MUID:94226599
 A:Accession: S43846
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-860 <BLA>
 A:Cross-references: GB:S71569; GB:X76919; NID:g560648; PIDN:AAB30669.1; PID:g560649
 C:Superfamily: Streptomyces endo-1,4-beta-xylanase A homology; fungal cellulose-binding
 F:49-325/Domain: Streptomyces endo-1,4-beta-xylanase A homology <SXY>
 F:829-860/Domain: fungal cellulose-binding domain homology <PCB>

Query Match 31.68; Score 61; DB 2; Length 860;
 Best Local Similarity 36.18; Pred. No. 6.8;
 Matches 13; Conservative 3; Mismatches 20; Indels 0; Gaps 0;

QY 2 EKLFEVKLELENKQSGGLGKAQCAALWLQCSGG 37
 |||||
 DB 801 QKVTVTVESQPTQGGNNGGNCARAKWGCGGNG 836

RESULT 3

S61429
 cystathionine beta-lyase (EC 4.4.1.8) precursor - Arabidopsis thaliana
 N:Alternate names: protein F2413.130
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 27-Apr-1996 #sequence_revision 27-Feb-1997 #text_change 19-May-2000
 C:Accession: S61429; T47765
 R:Ravanel, S.; Ruffet, M.L.; Douce, R.
 Plant Mol. Biol. 29, 875-882, 1995
 A:Title: Cloning of an Arabidopsis thaliana cDNA encoding cystathionine beta-lyase by fu
 A:Reference number: S61429; MUID:96128029
 A:Accession: S61429
 A:Molecule type: mRNA
 A:Residues: 1-464 <RAY>
 A:Cross-references: EMBL:L40511; NID:g704396; PIDN:AAA99176.1; PID:g704397
 R:Nyakatura, G.; Fartmann, B.; Dauner, D.; Stert, W.; Holland, R.; Weichselgartner, M.;
 submitted to the Protein Sequence Database, February 2000
 A:Reference number: Z24475
 A:Accession: T47765
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-464 <NVA>
 A:Cross-references: EMBL:AL138655
 A:Experimental source: cultivar Columbia; BAC clone F2413
 C:Genetics:
 A:Map position: 3
 A:Genome: nuclear
 A:Introns: 21/3; 78/1; 116/3; 140/2; 171/1; 200/2; 237/3; 270/1; 297/2; 336/3; 372/3; 40

C:Superfamily: O-succinylhomoserine (thiol)-lyase
 C:Keywords: carbon-sulfur lyase; chloroplast; methionine biosynthesis; phosphoprotein; F
 F:1-70/Domain: transit peptide (chloroplast) #status predicted <NTP>
 F:71-464/product: cystathionine beta-lyase #status predicted <MAT>
 F:278/Binding site: pyridoxal phosphate (lys) (covalent) #status predicted

Query Match 30.18; Score 58; DB 2; Length 464;
 Best Local Similarity 48.68; Pred. No. 9.3;
 Matches 17; Conservative 2; Mismatches 10; Indels 6; Gaps 3;

QY 2 EKLFEVKLELENKQSGGLGKAQCAALWLQCSGG 36
 |||||
 DB 296 EKLAKVYF-LQNSRSGSLAPPDC---WL-CLRG 324

RESULT 4

DB4607
 hypothetical protein At2g21970 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
 C:Accession: DB4607
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
 D.; Niernan, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
 Nature 402, 761-768, 1999
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A:Reference number: A84420; MUID:20083487
 A:Accession: DB4607
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-202 <STO>
 A:Cross-references: GB:AE002093; NID:g4417289; PIDN:AAD20414.1; GSPDB:GN00139
 C:Genetics:
 A:Gene: At2g21970
 A:Map position: 2

Query Match 29.08; Score 56; DB 2; Length 202;
 Best Local Similarity 43.38; Pred. No. 7.8;
 Matches 13; Conservative 5; Mismatches 10; Indels 2; Gaps 1;

QY 4 LFEVKLELENKQSGGLGKAQCAAL--WL 31
 |||||
 DB 130 LFKKLDVEGLSEAGLAGLAAMGCAAMFAWL 159

RESULT 5

JC7621
 capsacin receptor, VR1 - human
 C:Species: Homo sapiens (man)
 C:Date: 30-Jun-2001 #sequence_revision 30-Jun-2001 #text_change 30-Jun-2001
 C:Accession: JC7621
 R:Corrington, D.N.; Crandall, M.; Sanchez, J.F.; Zou, T.; Krause, J.E.; White, G.
 Biochem. Biophys. Res. Commun. 281, 1183-1189, 2001
 A:Title: The tissue distribution and functional characterization of human VR1.
 A:Reference number: JC7621; MUID:21139751; PMID:11243859
 A:Contents: Dorsal root ganglia
 A:Accession: JC7621
 A:Molecule type: mRNA
 A:Residues: 1-839 <COR>
 A:Cross-references: GB:AF196175
 C:Comment: This receptor, a transmembrane protein with many phosphorylation sites, is
 its integrative activation by several noxious stimuli, and plays an important role in

C:Genetics:
 A:Gene: vrl
 A:Map position: 17p13
 C:Keywords: transmembrane protein
 F:201-233/Domain: ankyrin #status predicted <ANK1>
 F:248-280/Domain: ankyrin #status predicted <ANK2>
 F:333-365/Domain: ankyrin #status predicted <ANK3>
 F:433-455/Domain: transmembrane #status predicted <TM1>
 F:477-495/Domain: transmembrane #status predicted <TM2>
 F:508-531/Domain: transmembrane #status predicted <TM3>
 F:543-569/Domain: transmembrane #status predicted <TM4>
 F:578-597/Domain: transmembrane #status predicted <TM5>
 F:624-644/Region: pore loop #status predicted
 F:656-684/Domain: transmembrane #status predicted <TM6>

Query Match 28.28; Score 54.5; DB 2; Length 839;
 Best Local Similarity 43.28; Pred. No. 47;
 Matches 16; Conservative 3; Mismatches 11; Indels 7; Gaps 1;

QY 3 KLFKEVKLELENKQSGGLGKAQCAALWLQ 32

Db 319 KLHPTKLEELTNKGMTPALAAAGTGKIGVLAVILQ 355
 || :||||| :|| :||| :|||

RESULT 6

T13636

probable minor structural protein - Streptococcus phage phi-Sf111

C:Species: Streptococcus phage phi-Sf111

C>Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 13-Aug-1999

C:Accession: T13636

R:Lucchini, S.; Desiere, F.; Bruessow, H.

Virology 246, 63-73, 1998

A:Title: The structural gene module in Streptococcus thermophilus bacteriophage phi Sf111

A:Reference number: 217696; MUID:98321150

A:Accession: T13636

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1000 <LUC>

A:Cross-references: EMBL:AF057033; NID:g3320432; PID:g3540286; PIDN:AAAC34413.1

A:Experimental source: specific_host Streptococcus thermophilus

Query Match 27.5%; Score 53; DB 2; Length 1000;

Best Local Similarity 40.0%; Pred. No. 86;

Matches 12; Conservative 3; Mismatches 13; Indels 2; Gaps 1;

QY 10 LEELENQKSGGLGKQAQCAAL--WLQCASGG 37
 ||| :||| :||| :||| :|||

Db 470 LNEMOGLKRRVNGQCQAALAAWYMKLGG 499
 ||| :||| :||| :||| :|||

RESULT 7

T13331

probable tail protein - Streptococcus phage phi-O1205

C:Species: Streptococcus phage phi-O1205

C>Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 24-Sep-1999

C:Accession: T13331

R:Stanley, E.; Fitzgerald, G.F.; Le Marrec, C.; Fayard, B.; van Sinderen, D.

Microbiology 143, 3417-3429, 1997

A:Title: Sequence analysis and characterization of phi O1205, a temperate bacteriophage

A:Reference number: 217654; MUID:98048466

A:Accession: T13331

A>Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1006 <STA>

A:Cross-references: EMBL:U88974; NID:g2444080; PID:g2444121; PIDN:AACT9557.1

A:Experimental source: host Streptococcus thermophilus strain CNR21205

Query Match 27.5%; Score 53; DB 2; Length 1006;

Best Local Similarity 40.0%; Pred. No. 87;

Matches 12; Conservative 3; Mismatches 13; Indels 2; Gaps 1;

QY 10 LEELENQKSGGLGKQAQCAAL--WLQCASGG 37
 ||| :||| :||| :||| :|||

Db 470 LNEMOGLKRRVNGQCQAALAAWYMKLGG 499
 ||| :||| :||| :||| :|||

RESULT 8

AE1220

ethanolamine utilization protein Eute homolog Imol165 [imported] - Listeria monocytogenes

C:Species: Listeria monocytogenes

C>Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001

C:Accession: AE1220

R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.

Science 294, 849-852, 2001

A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Maok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.

A:Title: Comparative genomics of Listeria species.

A:Reference number: AB1077; MUID:21537279; PMID:11679669

A:Accession: AE1220

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-469 <GLA>
 A:Cross-references: GB:NC_003210; PIDN:CAC99243.1; PID:g16410581; GSPDB:GN00177
 A:Experimental source: strain EGD-e
 C:Genetics:
 A:Gene: Imol165

Query Match 26.9%; Score 52; DB 2; Length 469;

Best Local Similarity 61.1%; Pred. No. 57;

Matches 11; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 MEKLFKEVKLELENQKQK 18
 :||| :||| :||| :|||

Db 6 LEKLVKVKLEKLAQKQK 23
 :||| :||| :||| :|||

RESULT 9

T05299

hypothetical protein F26P21.30 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 24-Nov-1999

C:Accession: T05299

R:Bevan, M.; Terry, N.; Ardiles, W.; Buysshaert, C.; Dasseville, R.; De Clerck, R.;

ewes, H.W.; Mayer, K.F.X.; Schueller, C.

submitted to the Protein Sequence Database, October 1998

A:Reference number: 215407

A:Accession: T05299

A:Molecule type: DNA

A:Residues: 1-761 <BEV>

A:Cross-references: EMBL:AL031804

A:Experimental source: cultivar Columbia; BAC clone F26P21

C:Genetics:

A:Map position: 4

A:Introns: 123/2; 167/2; 221/3; 262/3; 281/3; 348/3; 385/3; 398/2; 450/3; 484/3; 492/

A:Note: F26P21.30

C:Superfamily: Arabidopsis thaliana hypothetical protein F26P21.30

Query Match 26.9%; Score 52; DB 2; Length 761;

Best Local Similarity 50.0%; Pred. No. 90;

Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 19 SGLGKAQCAALWLQCA 34
 | : | : | : | : |

Db 288 SQITKLECSAFWVQCA 303
 | : | : | : | : |

RESULT 10

D90252

conserved hypothetical protein [imported] - Sulfolobus solfataricus

C:Species: Sulfolobus solfataricus

C>Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 15-Jun-2001

C:Accession: D90252

R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder

arrett, R.A.; Ragan, M.A.; Senses, C.W.; Van der Oost, J.

submitted to GenBank, April 2001

A:Description: Sulfolobus solfataricus complete genome.

A:Reference number: A99139

A:Accession: D90252

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-295 <KUR>

A:Cross-references: GB:AE005641; NID:g13814189; PIDN:AAK41275.1; GSPDB:GN00155

C:Genetics:

A:Gene: SSO1005

C:Superfamily: Pyrococcus horikoshii hypothetical protein PH1050

Query Match 26.7%; Score 51.5; DB 2; Length 295;

Best Local Similarity 54.2%; Pred. No. 43;

Matches 13; Conservative 6; Mismatches 4; Indels 1; Gaps 1;

QY 3 KLFKEVKLELENQKQK-SGLGKAQ 25
 | : | : | : | : |

Db 270 KYVEEKELELENRLGKSLKAK 293
 | : | : | : | : |

Search completed: October 24, 2002, 09:29:48
Job time : 11.0219 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 09:24:01 ; Search time 4.32117 Seconds
(without alignments)
331.536 Million cell updates/sec

Title: US-09-894-030-3_COPY_1_37

Perfect score: 193

Sequence: 1 MEKLFKEVLEENKSGSLGKAQCAALWLQASGG 37

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match	Length	DB	ID	Description
1	58	30.1	464	1	METC_ARATH	P53780	arabidopsis
2	53.5	27.7	181	1	Y06E_BPT4	P13311	bacterioph
3	53.5	27.7	516	1	HXR2_DROME	Q9nft7	grosophila
4	51.5	26.7	564	1	HEMA_TACKA	P19695	influenza a
5	51	26.4	638	1	SVT_AQUAE	O67583	aquifex aeo
6	50.5	26.2	898	1	C1Z1_HUMAN	Q9ulv3	homo sapien
7	49.5	25.6	451	1	2222_HUMAN	Q9uk12	homo sapien
8	49.5	25.6	563	1	HEMA_IAMAA	P26137	influenza a
9	49.5	25.6	564	1	HEMA_IADAI	P03443	influenza a
10	49.5	25.6	564	1	HEMA_IARUD	P19700	influenza a
11	49.5	25.6	564	1	HEMA_IASE2	P19701	influenza a
12	49.5	25.6	564	1	HEMA_IATKM	P19702	influenza a
13	49.5	25.6	568	1	HEMA_IAMAB	P26136	influenza a
14	49.5	25.6	754	1	ASPH_BOVIN	Q28056	bos taurus
15	49	25.4	259	1	PB1_CAMJE	P45678	campylobact
16	49	25.4	678	1	CMC1_HUMAN	O75746	homo sapien
17	49	25.4	1288	1	VAC1_HELJP	Q92kw5	helicobacte
18	49	25.4	1296	1	VAC1_HELJP	Q48247	helicobacte
19	48.5	25.1	111	1	Y0CD_BACSU	P45939	bacillus su
20	48.5	25.1	248	1	DJ1_DROME	O01352	drosophila
21	48.5	25.1	280	1	KIC3_XENLA	P05782	xenopus lae
22	48.5	25.1	534	1	NAB4_YEAST	Q99383	saccharomyc
23	48.5	25.1	560	1	NODU_AZOCA	Q07759	azorhizobiu
24	48.5	25.1	802	1	PMT2_SCHPO	O42832	schizosacch
25	48.5	25.1	1287	1	VAC2_HELJP	P45945	helicobacte
26	48.5	25.1	1290	1	VAC1_HELJP	P55981	helicobacte
27	48.5	25.1	1291	1	VAC4_HELJP	Q48258	helicobacte
28	48	24.9	104	1	Y164_ARCFU	O28415	archaeoglob
29	48	24.9	501	1	XYLB_LACPE	P21939	lactobacill
30	48	24.9	1043	1	RAG1_HUMAN	P15918	homo sapien
31	47.5	24.6	257	1	SHP_HUMAN	Q15466	homo sapien
32	47	24.4	293	1	YNP7_CAEEL	P34560	caenorhabdi
33	47	24.4	500	1	ERR2_HUMAN	O95718	homo sapien

34	47	24.4	639	1	V70K_PLRVI	P17519	potato leaf
35	47	24.4	639	1	V70K_PLRVW	P11622	potato leaf
36	47	24.4	1065	1	CARB_BACCL	P46537	bacillus ca
37	47	24.4	1065	1	VINC_HUMAN	P18206	homo sapien
38	47	24.4	1065	1	VINC_MOUSE	Q64727	mus musculu
39	47	24.4	1122	1	EXSC_ECOLI	P07648	escherichia
40	47	24.4	1691	1	YLH4_CAEEL	P34358	caenorhabdi
41	47	24.4	1938	1	MYS_AEQIR	P24733	aesquipecten
42	46.5	24.1	325	1	YRP1_CAEEL	Q10043	caenorhabdi
43	46.5	24.1	472	1	LI26_CAEEL	Q27355	caenorhabdi
44	46.5	24.1	570	1	HEMA_IAMC6	P26094	influenza a
45	46.5	24.1	570	1	HEMA_IAMC7	P26095	influenza a

ALIGNMENTS

RESULT 1
METC_ARATH STANDARD; PRT; 464 AA.
ID METC_ARATH
AC P53780;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Cystathionine beta-lyase, chloroplast precursor (EC 4.4.1.8) (CBL)
DE (Beta-cystathionase) (Cysteine lyase).
GN AT3G57050 OR F24I3.130.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, COLUMBIA;
RX MEDLINE=96128029; PubMed=8541513;
RA Ravanel S., Ruffet M.L., Douce R.;
RT "Cloning of an Arabidopsis thaliana cDNA encoding cystathionine beta-
lyase by functional complementation in Escherichia coli.";
RL Plant Mol. Biol. 29:875-882(1995).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, Columbia;
RX MEDLINE=21016720; PubMed=11130713;
RA Salanoubat M., Lemcke K., Rieger M., Ansoorge W., Unsel M.,
FArtmann B., Valle G., Bloecker H., Perez-Alonso M., Obermaier B.,
De Senne V., Boutry M., Grivell L.A., Mache R., Puigdomenech P.,
De Simone V., Choiane N., Artiguenave F., Robert C., Brottier P.,
Wincker P., Catolico L., Weissbach J., Saurin W., Quetier F.,
Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,
Wurmbach E., Drzonek H., Erfle H., Jordan N., Bangert S., Nyakatura G.,
Wiedelmann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,
Vezi A., D'Angelo M., Pallavicini A., Toppo S., Simonati B.,
Conrad A., Hornschke K., Kauer G., Loehner T.-H., Nordliek G.,
Reichert J., Scharfe M., Schoen O., Barges M., Terol J., Climent J.,
Navarro P., Collado C., Perez-Perez A., Ottenwaelder B., Duchemin D.,
Cooke R., Laudie M., Berger-Llauró C., Purnelle B., Masuy D.,
de Haan M., Maarse A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,
Monfort A., Argirou A., Flores M., Liguori R., Vitale D.,
Mannhaupt G., Haase D., Schoof H., Rudd S., Zaccaria P., Jenkins J.,
Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Shea T.P.,
Rooney T., Rizzo M., Walts A., Utterback T., Fujii C.Y., Aken S.,
Creasy T.H., Haas B., Maiti R., Wu D., Peterson J., Van Aken S.,
Pai G., Millsch J., Sellers P., Gill J.E., Feldblyum T.V.,
Preuss D., Lin X., Nierman W.C., Salzberg S.L., White O., Venter J.C.,
Fraser C.M., Kaneko T., Nakamura Y., Sato S., Kato T., Asamizu E.,
Sasomoto S., Kimura T., Idesawa K., Kawashima K., Kishida Y.,
Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
Nakayama S., Nakazaki N., Shinpo S., Takeuchi C., Wada T.,
Watanabe A., Yamada M., Yasuda M., Tabata S.;
RT "Sequence and analysis of chromosome 3 of the plant Arabidopsis
thaliana.";
RL Nature 408:820-822(2000).

RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 CC -|- CATALYTIC ACTIVITY: ATP + D-hexose = ADP + D-hexose 6-phosphate.
 CC -|- PATHWAY: FIRST STEP OF SEVERAL METABOLIC PATHWAYS.
 CC -|- SIMILARITY: BELONGS TO THE HEXOKINASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF257590; AAG22892.1; ALT_SEQ.
 DR EMBL: AF257591; AAG22894.1; ALT_SEQ.
 DR EMBL: AF257592; AAG22896.1; ALT_SEQ.
 DR EMBL: AF257593; AAG22898.1; ALT_SEQ.
 DR EMBL: AF257594; AAG22900.1; ALT_SEQ.
 DR EMBL: AF257595; AAG22902.1; ALT_SEQ.
 DR EMBL: AF257596; AAG22904.1; ALT_SEQ.
 DR EMBL: AF257597; AAG22906.1; ALT_SEQ.
 DR EMBL: AF257598; AAG22908.1; ALT_SEQ.
 DR EMBL: AF257599; AAG22910.1; ALT_SEQ.
 DR EMBL: AF257600; AAG22912.1; ALT_SEQ.
 DR EMBL: AF257601; AAG22914.1; ALT_SEQ.
 DR EMBL: AF257602; AAG22916.1; ALT_SEQ.
 DR EMBL: AF257603; AAG22918.1; ALT_SEQ.
 DR EMBL: AF257604; AAG22920.1; ALT_SEQ.
 DR EMBL: AF257605; AAG22922.1; ALT_SEQ.
 DR EMBL: AF257606; AAG22924.1; ALT_SEQ.
 DR EMBL: AF257607; AAG22926.1; ALT_SEQ.
 DR EMBL: AF257608; AAG22928.1; ALT_SEQ.
 DR EMBL: AJ271350; CAB67701.1; ALT_SEQ.
 DR EMBL: AJ271350; CAB72132.1; ALT_SEQ.
 DR EMBL: AE003756; AAF56591.1; ALT_SEQ.
 DR FlyBase; FBgn0042710; Hex-t2.
 DR InterPro; IPR001312; Hexokinase.
 DR Pfam; PF00349; hexokinase; 1.
 DR PRINTS; PR00475; HEXOKINASE.
 DR PRODOM; PD001109; hexokinase; 1.
 DR PROSITE; PS00378; HEXOKINASES; 1.
 DR Transferase; Kinase; Glycolysis; ATP-binding;
 DR Multigene family; Polymorphism.
 FT BINDING 158 158
 FT DOMAIN 197 223
 FT VARIANT 197 197
 FT VARIANT 224 224
 FT VARIANT 282 282
 FT VARIANT 410 410
 FT VARIANT 514 516
 FT SC96E3_12.3; HFL97E3_8; HFL97E3_12,
 FT HFL97E3_16; ZIM(S)E3_24 AND ZIM(S)E3_35).
 FT MISSING (IN REF. 1).
 FT CONFLICT 12 17
 SQ SEQUENCE 516 AA; 57122 MW; 01A0492BCDD567C CRC64;
 Query Match 27.7%; Score 53.5; DB 1; Length 516;
 Best Local Similarity 44.4%; Pred. No. 11;
 Matches 16; Conservative 5; Mismatches 14; Indels 1; Gaps 1;
 QY 1 MEKLFKE-VKLEENOKSGGLGKAAALWLQAS 35
 DB 479 MKKLLKPGVKELVSESGSRGALVAATAVQAKS 514
 RESULT 4

HEMA_IACKA
 ID HEMA_IACKA STANDARD; PRT; 564 AA.
 AC P19695;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
 DE Hemagglutinin HA2 chain].
 GN HA.
 OS Influenza A virus (strain A/Chicken/Alabama/1/75).
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 OC Influenza virus A and B group; Influenza A viruses;
 OC Influenza A virus.
 OX NCBI_TaxID=11338;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89204912; PubMed=2705304;
 RA Donis R.O., Bean W.J., Kawaoka Y., Webster R.G.;
 RT "Distinct lineages of influenza virus H4 hemagglutinin genes in
 RT different regions of the world.";
 RL Virology 169:408-417(1989).
 CC -|- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CC CELL RECEPTORS AND FOR INITIATING INFECTION.
 CC -|- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
 CC -|- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M25288; AAA43221.1;
 DR FIR; F34214; HMVFP6.
 DR HSP; P03437; 2HMG.
 DR InterPro; IPR001364; Hemagglutn.
 DR Pfam; PF00509; Hemagglutinin; 1.
 DR PRINTS; PR00329; HEMAGGLUTN12.
 DR PRODOM; PD000225; Hemagglutn; 1.
 DR Envelope protein; Hemagglutinin; Glycoprotein; Signal.
 FT SIGNAL 1 16
 FT CHAIN 17 342
 FT CHAIN 344 564
 FT CARBOHYD 18 18
 FT CARBOHYD 34 34
 FT CARBOHYD 178 178
 FT CARBOHYD 310 310
 FT CARBOHYD 497 497
 SQ SEQUENCE 564 AA; 63264 MW; B267C2A13DC1931C CRC64;
 Query Match 26.7%; Score 51.5; DB 1; Length 564;
 Best Local Similarity 36.4%; Pred. No. 22;
 Matches 12; Conservative 7; Mismatches 11; Indels 3; Gaps 1;
 QY 1 MEKLFKEVKLEENOKSGGLGKAAALWLQOC 33
 DB 458 MDKLFERVRQLRENAEDKNG---CFEIPHC 487
 RESULT 5
 SYT_AQUAE STANDARD; PRT; 638 AA.
 ID SYT_AQUAE
 AC O67583;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Threonyl-tRNA synthetase (EC 6.1.1.3) (Threonine--trna ligase)
 DE (ThRS).
 GN THRS OR AQ_1667.
 OS Aquifex aeolicus.


```
OC Influenza virus A and B group; Influenza A viruses;
OC Influenza A virus.
OX NCBI_TaxID=11346;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89204912; PubMed=2705304;
RA Denis R.O., Bean W.J., Kawaoka Y., Webster R.G.;
RT "Distinct lineages of influenza virus H4 hemagglutinin genes in
RL different regions of the world.";
RL Virology 169:408-417(1989).
[2]
RN SEQUENCE OF 1-101 FROM N.A.
RP MEDLINE=82150925; PubMed=6174976;
RX Air G.M.;
RT "Sequence relationships among the hemagglutinin genes of 12 subtypes
of influenza A virus.";
RL Proc. Natl. Acad. Sci. U.S.A. 78:7639-7643(1981).
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
(HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL; M25287; AAA43220.1; -
DR EMBL; J02102; AAA43179.1; -
DR PIR; E34214; HMIVF5.
DR PIR; A04054; HMIVF4.
DR HSP; P03437; 2HMG.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom: PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.
FT SIGNAL 1 16
FT CHAIN 17 342 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 344 564 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 18 18 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 34 34 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 178 178 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 310 310 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 497 497 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 69 69 R -> K (IN REF. 2).
SQ SEQUENCE 564 AA; 63381 MW; 808C4BBE9E15C94D CRC64;

Query Match 25.6%; Score 49.5; DB 1; Length 564;
Best Local Similarity 36.4%; Pred. No. 40;
Matches 12; Conservative 6; Mismatches 12; Indels 3; Gaps 1;

QY 1 MEKLFKEVKLEELNKGSLGKGAQCAALWLQC 33
| | | | : : : | | | : : |
Db 458 MNKLFERVRQLRENAEDKNG---CFEIPHCQ 487

RESULT 10
HEMA_IARUD STANDARD; PRT; 564 AA.
AC P19700;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain].
GN HA.
OS Influenza A virus (strain A/Ruddy Turnstone/New Jersey/47/85).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;

Query Match 25.6%; Score 49.5; DB 1; Length 564;
Best Local Similarity 36.4%; Pred. No. 40;
Matches 12; Conservative 6; Mismatches 12; Indels 3; Gaps 1;

QY 1 MEKLFKEVKLEELNKGSLGKGAQCAALWLQC 33
| | | | : : : | | | : : |
Db 458 MNKLFERVRQLRENAEDKNG---CFEIPHCQ 487

RESULT 11
HEMA_IASE2 STANDARD; PRT; 564 AA.
AC P19701;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain].
GN HA.
OS Influenza A virus (strain A/Seal/Massachusetts/133/82).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OC Influenza A virus.
OX NCBI_TaxID=11460;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89204912; PubMed=2705304;
RA Denis R.O., Bean W.J., Kawaoka Y., Webster R.G.;
RT "Distinct lineages of influenza virus H4 hemagglutinin genes in
RL different regions of the world.";
RL Virology 169:408-417(1989).
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
(HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
-----
DR HSP; P03437; 2HMG.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom: PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.
FT SIGNAL 1 16
FT CHAIN 17 342 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 344 564 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 18 18 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 34 34 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 178 178 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 310 310 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 497 497 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 69 69 R -> K (IN REF. 2).
SQ SEQUENCE 564 AA; 63271 MW; D4010C2548AE3EFC CRC64;
```

```
OC Influenza virus A and B group; Influenza A viruses;
OC Influenza A virus.
OX NCBI_TaxID=11458;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89204912; PubMed=2705304;
RA Denis R.O., Bean W.J., Kawaoka Y., Webster R.G.;
RT "Distinct lineages of influenza virus H4 hemagglutinin genes in
RL different regions of the world.";
RL Virology 169:408-417(1989).
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
(HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
-----
DR PIR; G34214; HMIVF7.
DR HSP; P03437; 2HMG.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom: PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.
FT SIGNAL 1 16
FT CHAIN 17 342 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 344 564 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 18 18 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 34 34 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 178 178 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 310 310 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 497 497 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 564 AA; 63271 MW; D4010C2548AE3EFC CRC64;

Query Match 25.6%; Score 49.5; DB 1; Length 564;
Best Local Similarity 36.4%; Pred. No. 40;
Matches 12; Conservative 6; Mismatches 12; Indels 3; Gaps 1;

QY 1 MEKLFKEVKLEELNKGSLGKGAQCAALWLQC 33
| | | | : : : | | | : : |
Db 458 MNKLFERVRQLRENAEDKNG---CFEIPHCQ 487

RESULT 11
HEMA_IASE2 STANDARD; PRT; 564 AA.
AC P19701;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain].
GN HA.
OS Influenza A virus (strain A/Seal/Massachusetts/133/82).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OC Influenza A virus.
OX NCBI_TaxID=11460;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89204912; PubMed=2705304;
RA Denis R.O., Bean W.J., Kawaoka Y., Webster R.G.;
RT "Distinct lineages of influenza virus H4 hemagglutinin genes in
RL different regions of the world.";
RL Virology 169:408-417(1989).
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
(HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
-----
DR HSP; P03437; 2HMG.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
```

ProDom: PD000225; Hemagglutn: 1.
 KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.
 FT SIGNAL 1 16
 FT CHAIN 17 342 HEMAGGLUTININ HA1 CHAIN.
 FT CHAIN 344 564 HEMAGGLUTININ HA2 CHAIN.
 FT CARBOHYD 18 18 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 34 34 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 178 178 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 310 310 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 497 497 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 564 AA; 63165 MW; 455D17BIADD2B82B CRC64;
 Query Match 25.6%; Score 49.5; DB 1; Length 564;
 Best Local Similarity 36.4%; Pred. No. 40;
 Matches 12; Conservative 6; Mismatches 12; Indels 3; Gaps 1;
 QY 1 MEKLFKEVKLEENQKSGGLGKAQCAALWLQC 33
 DB 458 MNKLFERVRRLRENAEDKNG---CFEIPHQC 487
 RESULT 12
 HEMA_IATKM
 ID HEMA_IATKM STANDARD; PRT; 564 AA.
 AC P19702;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
 DE Hemagglutinin HA2 chain].
 OS Influenza A virus (strain A/Turkey/Minnesota/833/80).
 OS Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 OC Influenza virus A and B group; Influenza A viruses;
 OC Influenza A virus.
 OX NCBI_TaxID=11468;
 RN SEQUENCE FROM N.A.
 RX MEDLINE=89204912; PubMed=27053304;
 RA Donis R.O., Bean W.J., Kawoka Y., Webster R.G.;
 RT "Distinct lineages of influenza virus H4 hemagglutinin genes in
 RT different regions of the world.";
 RL Virology 169:408-417(1989).
 CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CC CELL RECEPTORS AND FOR INITIATING INFECTION.
 CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
 CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M25290; AAA3223.1; .
 CC PIR; H32214; HMIIVF6.
 CC HSP; P03437; 2HMG.
 CC InterPro; IPR001364; Hemagglutn.
 CC Pfam; PF00509; Hemagglutinin; 1.
 CC PRINTS; PR00329; HEMAGGLUTN12.
 CC ProDom; PD000225; Hemagglutn: 1.
 KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.
 FT SIGNAL 1 16
 FT CHAIN 17 342 HEMAGGLUTININ HA1 CHAIN.
 FT CHAIN 344 564 HEMAGGLUTININ HA2 CHAIN.
 FT CARBOHYD 18 18 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 34 34 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 178 178 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 310 310 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 497 497 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 497 497 N-LINKED (GLCNAC. . .) (POTENTIAL).

SQ SEQUENCE 564 AA; 63249 MW; A8A3D998BBE3A58F CRC64;
 Query Match 25.6%; Score 49.5; DB 1; Length 564;
 Best Local Similarity 36.4%; Pred. No. 40;
 Matches 12; Conservative 6; Mismatches 12; Indels 3; Gaps 1;
 QY 1 MEKLFKEVKLEENQKSGGLGKAQCAALWLQC 33
 DB 458 MNKLFERVRRLRENAEDKNG---CFEIPHQC 487
 RESULT 13
 HEMA_IAMAB
 ID HEMA_IAMAB STANDARD; PRT; 568 AA.
 AC P26136;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
 DE Hemagglutinin HA2 chain].
 OS Influenza A virus (strain A/Mallard/Astrakhan/263/82) (Influenza A
 OS virus (strain A/Mallard/Gurjev/263/82)).
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 OC Influenza virus A and B group; Influenza A viruses;
 OC Influenza A virus.
 OX NCBI_TaxID=11434;
 RN SEQUENCE FROM N.A.
 RX MEDLINE=9104942; PubMed=2238469;
 RA Kawoka Y., Yarnikova S., Chambers T.M., Lvov D.K., Webster R.G.;
 RT "Molecular characterization of a new hemagglutinin, subtype H14, of
 RT influenza A virus.";
 RL Virology 179:759-767(1990).
 CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CC CELL RECEPTORS AND FOR INITIATING INFECTION.
 CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
 CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M35997; .; NOT_ANNOTATED_CDS.
 CC PIR; A46339; A46339.
 CC HSP; P03437; IHTM.
 CC InterPro; IPR001364; Hemagglutn.
 CC Pfam; PF00509; Hemagglutinin; 1.
 CC PRINTS; PR00329; HEMAGGLUTN12.
 CC ProDom; PD000225; Hemagglutn: 1.
 KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.
 FT SIGNAL 1 17
 FT CHAIN 18 346 HEMAGGLUTININ HA1 CHAIN.
 FT CHAIN 348 568 HEMAGGLUTININ HA2 CHAIN.
 FT CARBOHYD 21 21 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 62 62 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 182 182 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 242 242 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 314 314 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 501 501 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 568 AA; 63149 MW; 34D2F5444BEF6BF9 CRC64;
 Query Match 25.6%; Score 49.5; DB 1; Length 568;
 Best Local Similarity 36.4%; Pred. No. 41;
 Matches 12; Conservative 6; Mismatches 12; Indels 3; Gaps 1;
 QY 1 MEKLFKEVKLEENQKSGGLGKAQCAALWLQC 33

[illegible][illegible]

DR SMART; SM00062; PBPB; 1.
DR PROSITE; PS01039; SBP_BACTERIAL_3; 1.
KW Signal; Transport; Antigen; Complete proteome.
FT SIGNAL 1 26
FT CHAIN 27 259 MAJOR CELL-BINDING FACTOR.
SQ SEQUENCE 259 AA; 28177 MW; DA132ECE87BBCBD5 CRC64;
Query Match 25.4%; Score 49; DB 1; Length 259;
Best Local Similarity 41.7%; Pred. NO. 23;
Matches 10; Conservative 5; Mismatches 9; Indels 0; Gaps 0;
Qy 4 LFKEVKLEELNKGSGLGKQAQA 27
Db 134 VLREKYKSLADMKGANIGVAQAA 157

Search completed: October 24, 2002, 09:28:23
Job time : 7.32117 secs

GenCore version 5.1.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2002, 09:24:21 ; Search time 11.8832 Seconds
(without alignments)
538.643 Million cell updates/sec

Title: US-09-894-030-3_COPY_1_37

Perfect score: 193

Sequence: 1 MEKLFKEVKLEELNKGSGLGKRAQCAALWLCASGG 37

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL19.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description
1	193	100.0	56	9	064033	O64033 bacterioph
2	193	100.0	56	16	034781	O34781 bacillus su
3	61	31.6	860	3	Q02290	Q02290 neocallimas
4	56	29.0	170	10	Q94F34	Q94F34 arabidopsis
5	56	29.0	202	10	Q9SJ02	Q9SJ02 arabidopsis
6	56	29.0	513	5	Q9N2W3	Q9N2W3 caenorhabdi
7	54.5	28.2	453	5	Q9GN24	Q9GN24 drosophila
8	54.5	28.2	453	5	Q9GN13	Q9GN13 drosophila
9	54.5	28.2	453	5	Q9GN05	Q9GN05 drosophila
10	54.5	28.2	511	4	Q9H303	Q9H303 homo sapien
11	54.5	28.2	839	4	Q9H304	Q9H304 homo sapien
12	54.5	28.2	839	4	Q9NY22	Q9NY22 homo sapien
13	54.5	28.2	839	4	Q9NQ74	Q9NQ74 homo sapien
14	54.5	28.2	839	4	Q9H0G9	Q9H0G9 homo sapien
15	54	28.0	634	15	Q9IEZ2	Q9IEZ2 caprine art
16	53.5	27.7	453	5	Q95U08	Q95U08 drosophila

17	53	27.5	1000	9	080181	080181 streptococc
18	53	27.5	1006	9	034073	034073 streptococc
19	52.5	27.2	729	2	Q92487	Q92487 aeromonas p
20	52.5	27.2	1322	11	Q9Q2P6	Q9Q2P6 mus musculu
21	52	26.9	383	10	Q9LWJ2	Q9LWJ2 oryza sativ
22	52	26.9	475	10	Q9LWJ5	Q9LWJ5 oryza sativ
23	52	26.9	713	10	Q9ASW9	Q9ASW9 arabidopsis
24	52	26.9	761	10	082634	082634 arabidopsis
25	52	26.9	1287	2	Q93UM8	Q93UM8 helicobacte
26	52	26.9	1288	2	Q93UM9	Q93UM9 helicobacte
27	52	26.9	1289	2	Q93UM6	Q93UM6 helicobacte
28	51.5	26.7	295	17	Q97ZB8	Q97ZB8 sulfolobus
29	51.5	26.7	443	10	Q93XB2	Q93XB2 rosa hybrid
30	51.5	26.7	453	5	Q9GTU1	Q9GTU1 drosophila
31	51.5	26.7	811	10	Q9FWG8	Q9FWG8 oryza sativ
32	51.5	26.7	1024	10	Q9FXI1	Q9FXI1 arabidopsis
33	51.5	26.7	2584	5	Q9TW88	Q9TW88 caenorhabdi
34	51.5	26.7	2606	5	021920	021920 caenorhabdi
35	51	26.4	361	16	Q9IIC0	Q9IIC0 pseudomonas
36	51	26.4	368	2	Q9L4D6	Q9L4D6 xanthomonas
37	51	26.4	514	2	Q937U6	Q937U6 borrelia he
38	51	26.4	526	10	Q9LMD4	Q9LMD4 arabidopsis
39	51	26.4	533	10	Q9SSG2	Q9SSG2 arabidopsis
40	51	26.4	558	3	Q9UUA3	Q9UUA3 schizosacch
41	51	26.4	1738	10	Q9SI41	Q9SI41 arabidopsis
42	50.5	26.2	633	3	059795	059795 schizosacch
43	50.5	26.2	837	4	Q9BTG3	Q9BTG3 homo sapien
44	50.5	26.2	842	4	Q9H868	Q9H868 homo sapien
45	50	25.9	148	17	059083	059083 pyrococcus

ALIGNMENTS

RESULT 1

O64033 ID O64033 PRELIMINARY; PRT; 56 AA.
AC O64033;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PUTATIVE LIPOPROTEIN.
GN YOLG.
OS Bacteriophage SPBC2.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.
OX NCBI_TaxID=66797;
RN [1]
RP SEQUENCE FROM N.A.
RA Lazarevic V., Duesterhoeft A., Soldo B., Hilbert H., Mael C.,
RA Karamata D.;
RT "The complete nucleotide sequence of the Bacillus subtilis Spbetac2
RT prophage."
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF020713; AAC12992.1; -
KW Lipoprotein.
SQ SEQUENCE 56 AA; 5982 MW; 79EC0BF822F9F4C0 CRC64;

Query Match 100.0%; Score 193; DB 9; Length 56;

Best Local Similarity 100.0%; Pred. No. 2.4e-19;

Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEKLFKEVKLEELNKGSGLGKRAQCAALWLCASGG 37

|||||

DB 1 MEKLFKEVKLEELNKGSGLGKRAQCAALWLCASGG 37

RESULT 2

O34781 ID O34781 PRELIMINARY; PRT; 56 AA.
AC O34781;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)

FT	REPEAT	639	646	39.
FT	REPEAT	647	654	40.
FT	REPEAT	655	662	41.
FT	REPEAT	663	670	42.
FT	REPEAT	671	678	43.
FT	REPEAT	679	686	44.
FT	REPEAT	687	694	45.
FT	REPEAT	695	702	46.
FT	REPEAT	703	710	47.
FT	REPEAT	711	718	48.
FT	REPEAT	719	726	49.
FT	REPEAT	727	734	50.
FT	REPEAT	735	742	51.
FT	REPEAT	743	750	52.
FT	REPEAT	751	758	53.
FT	REPEAT	759	766	54.
FT	REPEAT	767	774	55.
FT	REPEAT	775	782	56.
FT	REPEAT	783	790	57.
FT	DOMAIN	832	859	
FT	DLSULFID	278	284	
FT	CARBOHYD	295	295	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	309	309	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	359	359	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	374	374	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	390	390	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	406	406	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	845	845	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	SEQUENCE	860 AA;	88052 MW;	677F9032F0A87085 CRC64;
QY	Query Match	31.6%;	Score 61;	DB 3; Length 860;
Db	Best Local Similarity	36.1%;	Pred. No. 7.2;	
	Matches 13;	Conservative 3;	Mismatches 20;	Indels 0; Gaps 0;
QY	2 EKLPEKYLELENQKSGLGKAKOCAAALWLOQASGG 37	:	:	:
Db	801 QKVTVTVTVESQPTQGMNMGGNCAARKWGQCNGNG 836	:	:	:
RESULT 4				
Q94F34				
ID	Q94F34	PRELIMINARY;	PRT;	170 AA.
AC	Q94F34;			
DT	01-DEC-2001 (TrEMBLrel. 19, Created)			
DT	01-DEC-2001 (TrEMBLrel. 19, Last sequence update)			
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)			
DE	UNKNOWN PROTEIN.			
OS	ATG2G21970.			
GN	Arabidopsis thaliana (Mouse-ear cress).			
OC	Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;			
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsids.			
NCBI_Taxid=3702;				
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Lam B., Southwick A., Karlin-Neumann G., Nguyen M., Miranda M.,			
RA	Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,			
RA	Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,			
RA	Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,			
RA	Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,			
RA	Ecker J., Theologis A., Davis R.W.;			
RL	Submitted (May-2001) to the EMBL/GenBank/DDBJ databases.			
DR	EMBL; AF386927; AAK62372.1; -.			
SO	SEQUENCE 170 AA; 18220 MW; BD0FFB28DC3C01ED CRC64;			

Query Match	29.08;	Score 56;	DB 10;	Length 170;
Best Local Similarity	43.3%;	Pred. No. 6.3;		
Matches 13;	Conservative 5;	Mismatches 10;	Indels 2;	Gaps 1;
QY	4	LFKEVKLEENKSGSLGKAOCAAL--WL 31		
Db	98	LFKKLDVGLSEATGAGLAANGCAAMFAWL 127		

```

RESULT 5
Q9SJO2 PRELIMINARY; PRT; 202 AA.
AC Q9SJO2;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE ATG221970 PROTEIN (STRESS ENHANCED PROTEIN 2).
GN ATG221970 OR SEP2.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
[1]
SEQUENCE FROM N.A.
RP STRAIN=CV. COLUMBIA;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA Buell L.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.
RA Cronin L.A., Shen M., VanAken S.E., Unayam L., Tallon L.J., Gill J.E.
RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eissen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.;
RT "Sequence and analysis of chromosome II of Arabidopsis thaliana.";
RL Nature 402:761-768(1999).
[2]
SEQUENCE FROM N.A.
RP MEDLINE=2020706; PubMed=10725357;
RA Heddad M., Adamska I.;
RT "Light stress-regulated two-helix proteins in Arabidopsis thaliana
RT related to the chlorophyll a/b-binding gene family.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:3741-3746(2000).
DR EMBL; AC007019; AAD20414.1; -.
DR EMBL; AF137117; AAF61626.1; -.
SQ SEQUENCE 202 AA; 21989 MW; D300A5B1EE70F6F8 CRC64;

Query Match 29.08; Score 56; DB 10; Length 202;
Best Local Similarity 43.3%; Pred. No. 7.6;
Matches 13; Conservative 5; Mismatches 10; Indels 2; Gaps

QY 4 LFKVEKLELENKGSGLGKQAQAL--WL 31
|||:::| | | | | | | | | | | | | |
Db 130 LFKKLDVGLSEAGLAGLAAGCAAMFAWL 159

RESULT 6
Q9N2W3 PRELIMINARY; PRT; 513 AA.
AC Q9N2W3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 59.2 KDA PROTEIN.
GN Y94H6A.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
[1]
SEQUENCE FROM N.A.
RP STRAIN=BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
[2]
SEQUENCE FROM N.A.
RP STRAIN=BRISTOL N2;
RC Wollam C., Zidanic M., Du H.;

```

```
RT "The sequence of C. elegans cosmid Y94H6A.";
RN Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR EMBL; AC024876; AAF60896.1; -.
DR HSSP; P20393; 1A6Y.
DR InterPro; IPR000536; Hormone_rec_llg.
DR InterPro; IPR001628; zf-C4.
DR InterPro; IPR000130; Zn_Mtpeptdse.
DR Pfam; PF00104; hormone_rec; 1.
DR Pfam; PF00105; zf-C4; 1.
DR PRINTS; PR00047; STROIDFINGER.
DR SMART; SM00430; HOLI; 1.
DR SMART; SM00399; Znf_C4; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
KW DNA-binding; Hypothetical protein; Nuclear protein; Receptor;
KW Transcription regulation; Zinc-finger.
SQ SEQUENCE 513 AA; 59164 MW; C79EDF2C65E99A6E CRC64;

Query Match 29.0%; Score 56; DB 5; Length 513;
Best Local Similarity 40.8%; Pred. No. 20;
Matches 13; Conservative 4; Mismatches 15; Indels 0; Gaps 0;

QY 5 FKEVKLEENKGGSLGKACAAALWLQASG 36
| : : : | | | | | : : : | | | | |
Db 311 FLAIQREFFENCPCDGOIPACSAEWFQELSG 342

RESULT 7
Q9GN24
ID Q9GN24 PRELIMINARY; PRT; 453 AA.
AC Q9GN24;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE HEXOKINASE-T2.
OS Drosophila simulans (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7240;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HFL97_4S, VA96_7S, GA96_2S, AND GA96_4S;
RA Duvernell D.D., Eanes W.F.;
RT "Contrasting molecular population genetics of four hexokinases in
Drosophila melanogaster and Drosophila simulans.";
RL Genetics 0:0-0(2000).
DR EMBL; AF257649; AAG22968.1; -.
DR EMBL; AF257642; AAG22954.1; -.
DR EMBL; AF257644; AAG22958.1; -.
DR EMBL; AF257645; AAG22960.1; -.
DR HSSP; P19367; 1HKC.
DR FlyBase; FBgn0042855; Dsim\Hex-t2.
DR InterPro; IPR001312; Hexokinase.
DR Pfam; PF00349; hexokinase; 1.
DR PRINTS; PR00475; HEXOKINASE.
DR ProDom; PD001109; Hexokinase; 1.
DR PROSITE; PS00378; HEXOKINASES; 1.
KW Kinase.
SQ SEQUENCE 453 AA; 49856 MW; E530FBF0E717B821 CRC64;

Query Match 28.2%; Score 54.5; DB 5; Length 453;
Best Local Similarity 44.4%; Pred. No. 29;
Matches 16; Conservative 5; Mismatches 14; Indels 1; Gaps 1;

QY 1 MEKLFKE-VKLEENKGGSLGKACAAALWLQAS 35
| : : : | | | | | : : : | | | | |
Db 416 MKKLLKPGVKFELIVSDGSGRGAALVAATAVQAKS 451

RESULT 9
Q9GN05
ID Q9GN05 PRELIMINARY; PRT; 453 AA.
AC Q9GN05;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE HEXOKINASE-T2.
OS Drosophila simulans (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7240;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GA96_5S, DPF96_3S, CT96_2S, CT96_5S, AND VA96_5S;
RA Duvernell D.D., Eanes W.F.;
RT "Contrasting molecular population genetics of four hexokinases in
Drosophila melanogaster and Drosophila simulans.";
RL Genetics 0:0-0(2000).
DR EMBL; AF257646; AAG22962.1; -.
DR EMBL; AF257637; AAG22944.1; -.
DR EMBL; AF257638; AAG22946.1; -.
DR EMBL; AF257639; AAG22948.1; -.
DR EMBL; AF257640; AAG22950.1; -.
DR EMBL; AF257641; AAG22952.1; -.
DR HSSP; P19367; 1HKC.
SQ SEQUENCE 453 AA; 49843 MW; EB93A6537F1B6504 CRC64;

Query Match 28.2%; Score 54.5; DB 5; Length 453;
Best Local Similarity 44.4%; Pred. No. 29;
Matches 16; Conservative 5; Mismatches 14; Indels 1; Gaps 1;

QY 1 MEKLFKE-VKLEENKGGSLGKACAAALWLQAS 35
| : : : | | | | | : : : | | | | |
Db 416 MKKLLKPGVKFELIVSDGSGRGAALVAATAVQAKS 451

RESULT 9
Q9GN05
ID Q9GN05 PRELIMINARY; PRT; 453 AA.
AC Q9GN05;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE HEXOKINASE-T2.
OS Drosophila simulans (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7240;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GA96_5S, DPF96_3S, CT96_2S, CT96_5S, AND VA96_5S;
RA Duvernell D.D., Eanes W.F.;
RT "Contrasting molecular population genetics of four hexokinases in
Drosophila melanogaster and Drosophila simulans.";
RL Genetics 0:0-0(2000).
DR EMBL; AF257646; AAG22962.1; -.
DR EMBL; AF257637; AAG22944.1; -.
DR EMBL; AF257638; AAG22946.1; -.
DR EMBL; AF257639; AAG22948.1; -.
DR EMBL; AF257640; AAG22950.1; -.
DR EMBL; AF257641; AAG22952.1; -.
DR HSSP; P19367; 1HKC.
SQ SEQUENCE 453 AA; 49843 MW; EB93A6537F1B6504 CRC64;
```

DR FlyBase; FBgn0042855; Dsim\Hex-t2.
 DR InterPro; IPR001312; Hexokinase.
 DR Pfam; PF00349; hexokinase; 1.
 DR PRINTS; PR00475; HEXOKINASE.
 DR ProDom; PD001109; Hexokinase; 1.
 DR PROSITE; PS00378; HEXOKINASES; 1.
 KW Kinase.
 SQ SEQUENCE 453 AA; 49869 MW; FE87E3537F1B651F CRC64;

Query Match 28.2%; Score 54.5; DB 5; Length 453;
 Best Local Similarity 44.4%; Pred. No. 29;
 Matches 16; Conservative 5; Mismatches 14; Indels 1; Gaps 1;

QY 1 MEKLFKE-VKLELEENQKSGLGKAOCAALWLQAS 35
 ||| ||| | : : : ||| | ||| : | |
 DB 416 MKKLLKPGVKFELIVSEDSGSGAALVAVTAQVAKS 451

RESULT 10
 Q9H303 PRELIMINARY; PRT; 511 AA.
 AC Q9H303;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE CAPSAICIN RECEPTOR VARIANT.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.

RA Cortright D.N., Peck A., Zou T.;
 RT "Cloning and characterization of the human capsaicin receptor, VR1.";
 RL Submitted (Oct-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF196176; AAG43467.1; -.
 DR InterPro; IPR002110; ANK.
 DR SMART; SM00248; ANK; 4.
 DR PROSITE; PS50088; ANK_REPEAT; 1.
 DR PROSITE; PS50297; ANK_REPEAT; Repeat.
 KW ANK repeat; Receptor; Repeat.
 SQ SEQUENCE 511 AA; 56977 MW; 27BAFACEBBCE0946 CRC64;

Query Match 28.2%; Score 54.5; DB 4; Length 511;
 Best Local Similarity 43.2%; Pred. No. 33;
 Matches 16; Conservative 3; Mismatches 11; Indels 7; Gaps 1;

QY 3 KLFKEVKLELEENQKG-----SGLGKAOCAALWLQ 32
 || :||| ||| :|| :||| | ||
 DB 319 KLHPTLKLEELTNKGMTPLAALAAAGTGKIGVLAYILQ 355

RESULT 11
 Q9H304 PRELIMINARY; PRT; 839 AA.
 AC Q9H304;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE CAPSAICIN RECEPTOR.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.

RA Cortright D.N., Peck A., Zou T.;
 RT "Cloning and characterization of the human capsaicin receptor, VR1.";
 RL Submitted (Oct-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF196175; AAG43466.1; -.
 DR InterPro; IPR002110; ANK.
 DR InterPro; IPR000636; Cation_chan_non_lig.
 DR InterPro; IPR002111; Cat_channel_TrpL.

DR Pfam; PF00520; ion_trans; 1.
 DR SMART; SM00248; ANK; 4.
 DR PROSITE; PS50088; ANK_REPEAT; 1.
 DR PROSITE; PS50297; ANK_REPEAT; Repeat.
 KW ANK repeat; Receptor; Repeat.
 SQ SEQUENCE 839 AA; 94968 MW; EA9BF7D6023EF9F4 CRC64;

Query Match 28.2%; Score 54.5; DB 4; Length 839;
 Best Local Similarity 43.2%; Pred. No. 56;
 Matches 16; Conservative 3; Mismatches 11; Indels 7; Gaps 1;

QY 3 KLFKEVKLELEENQKG-----SGLGKAOCAALWLQ 32
 || :||| ||| :|| :||| | ||
 DB 319 KLHPTLKLEELTNKGMTPLAALAAAGTGKIGVLAYILQ 355

RESULT 12
 Q9NY22 PRELIMINARY; PRT; 839 AA.
 AC Q9NY22;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE VANILLOID RECEPTOR 1.
 GN VR1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.

RC TISSUE=DORSAL ROOT GANGLION;
 RX MEDLINE=21124568; PubMed=11226139;
 RA McIntyre P., McIachle L., Chambers A., Phillips E., Clarke M.,
 RA Savidge J., Peacock M., Shah K., Winter J., Weerasekera N., Webb M.,
 RA Rang H., Bevan S., James I.;
 RT "Pharmacological differences between human and rat vanilloid receptor
 1 (VR1).";
 RL Br. J. Pharmacol. 132:1084-1094(2001).
 DR EMBL; AJ272063; CAB89866.2; -.
 DR InterPro; IPR002110; ANK.
 DR InterPro; IPR000636; Cation_chan_non_lig.
 DR InterPro; IPR002111; Cat_channel_TrpL.
 DR Pfam; PF00023; ank; 3.
 DR Pfam; PF00520; ion_trans; 1.
 DR SMART; SM00248; ANK; 4.
 DR PROSITE; PS50088; ANK_REPEAT; 1.
 DR PROSITE; PS50297; ANK_REPEAT; Repeat.
 KW ANK repeat; Receptor; Repeat.
 SQ SEQUENCE 839 AA; 94968 MW; 773BF59D43189968 CRC64;

Query Match 28.2%; Score 54.5; DB 4; Length 839;
 Best Local Similarity 43.2%; Pred. No. 56;
 Matches 16; Conservative 3; Mismatches 11; Indels 7; Gaps 1;

QY 3 KLFKEVKLELEENQKG-----SGLGKAOCAALWLQ 32
 || :||| ||| :|| :||| | ||
 DB 319 KLHPTLKLEELTNKGMTPLAALAAAGTGKIGVLAYILQ 355

RESULT 13
 Q9NO74 PRELIMINARY; PRT; 839 AA.
 AC Q9NO74;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE VANILLOID RECEPTOR 1.
 GN VR1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;

```

RN RP SEQUENCE FROM N.A.
RA Kelsell R.E.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN RP SEQUENCE FROM N.A.
RX MEDLINE-20504379; PubMed-11050376;
RA Hayes P., Meadows H.J., Gunthorpe M., Harries M.H., Duckworth M.D.,
RA Cairns W., Harrison D.C., Clarke C., Ellington K., Prinjha R.K.,
RA Barton A.J., Medhurst A.D., Smith G.D., Topp S., Murdoch P.,
RA Sanger G.J., Terrett J., Jenkins O., Benham C.D., Randall A.D.,
RA Gloger I.S., Davis J.B.;
RT "Cloning and functional expression of a human orthologue of rat
RL vanilloid receptor-1.";
RL Pain 88:205-215(2000).
DR EMBL; AJ277028; CAB95729.1; -.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR000636; Cation_chan_non_lig.
DR InterPro; IPR002111; Cat_channel_trpL.
DR Pfam; PF00520; ion_trans; 1.
DR SMART; SM00248; ANK; 3.
DR PROSITE; PS50088; ANK_REPEAT; 1.
DR PROSITE; PS50297; ANK_REPEAT; 1.
DR ANK repeat; Receptor; Repeat.
KW ANK repeat; Receptor; Repeat.
SQ SEQUENCE 839 AA; 94924 MW; 7142FFAE43189ECC CRC64;

Query Match 28.2%; Score 54.5; DB 4; Length 839;
Best Local Similarity 43.2%; Pred. No. 56;
Matches 16; Conservative 3; Mismatches 11; Indels 7; Gaps 1;

QY 3 KLFKEVKLELENQKG-----SGLGKAQCAALWLO 32
DB 319 KLHPTLKEELTNKKGMTPLAALAAAGTGKIGVAYILQ 355

RESULT 14
Q9H0G9 PRELIMINARY; PRT; 839 AA.
AC Q9H0G9;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOPHETICAL 94.9 KDA PROTEIN.
GN DKF2P434K0220.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RX MEDLINE-21154917; PubMed-11230166;
RA Wiemann S., Weil B., Wellenreuther R., Gassenhuber J., Glassl S.,
RA Ansgorge W., Boecker M., Bloeker H., Bauersachs S., Blum H.,
RA Lauber J., Duesterhoeft A., Beyer A., Koehrer K., Strack N.,
RA Mewes H.W., Ottenwaelder B., Obermaier B., Tampe J., Heubner D.,
RA Wambutt R., Korn B., Klein M., Poustka A.;
RT "Towards a Catalog of Human Genes and Proteins: Sequencing and
RT Analysis of 500 Novel Complete Protein Coding Human cDNAs.";
RL Genome Res. 11:422-435(2001).
DR EMBL; AL136801; CAB66735.1; -.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR000636; Cation_chan_non_lig.
DR InterPro; IPR002111; Cat_channel_trpL.
DR Pfam; PF00520; ion_trans; 1.
DR SMART; SM00248; ANK; 4.
DR PROSITE; PS50088; ANK_REPEAT; 1.
DR PROSITE; PS50297; ANK_REPEAT; 1.
KW ANK repeat; Hypothetical protein; Repeat.
SQ SEQUENCE 839 AA; 94950 MW; EA9BF7D603AE4760 CRC64;

Query Match 28.2%; Score 54.5; DB 4; Length 839;
Best Local Similarity 43.2%; Pred. No. 56;

Matches 16; Conservative 3; Mismatches 11; Indels 7; Gaps 1;

QY 3 KLFKEVKLELENQKG-----SGLGKAQCAALWLO 32
DB 319 KLHPTLKEELTNKKGMTPLAALAAAGTGKIGVAYILQ 355

RESULT 15
Q9IEZ2 PRELIMINARY; PRT; 634 AA.
AC Q9IEZ2;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE SURFACE GLYCOPROTEIN PRECURSOR (FRAGMENT).
GN ENV.
OS Caprine arthritis encephalitis virus (CAEV).
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11660;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=021;
RX MEDLINE-20305014; PubMed-10846103;
RA Valas S., Benoit C., Baudry C., Perrin G., Mamoun R.Z.;
RT "Variability and immunogenicity of Caprine arthritis-encephalitis
RT virus surface glycoprotein.";
RL J. Virol. 74:6178-6185(2000).
DR EMBL; AJ400719; CAB95712.1; -.
KW Signal.
FT SIGNAL. 1 86 POTENTIAL.
FT CHAIN 87 >634 SURFACE GLYCOPROTEIN.
FT NON_TER 634 634
SQ SEQUENCE 634 AA; 72847 MW; 75B196E23560F23E CRC64;

Query Match 28.0%; Score 54; DB 15; Length 634;
Best Local Similarity 50.0%; Pred. No. 48;
Matches 12; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 2 EKLFEVKLELENQKSGGLGKRAQ 25
DB 13 EKNWKEVIIIEEEKKGCGTRRQ 36

Search completed: October 24, 2002, 09:29:16
Job time :15.8832 secs
```